

English (/complete-guide) عربي (/ar/complete-guide) Français (/fr/complete-guide)
русский (/ru/complete-guide) Español (/es/complete-guide)


Search...



HOME (/) ABOUT▼ TOPICS (/TOPICS)
INTERVENTIONS (/INTERVENTIONS)
SUBMIT AN INTERVENTION (/SUBMIT-INTERVENTION)

Home (/) > Complete Guide

COMPLETE GUIDE

 Printer Friendly, PDF & Email (<https://www.printfriendly.com/print?url=http://www.counteringdisinformation.org/node/2698>)

INTRODUCTION TO THE GUIDE

The online world plays an increasingly dominant role in shaping the public conversation and driving political events. Concurrently, disinformation, hate speech, and online extremism have seemingly saturated content on social media platforms, their harms compounded by ever more powerful network effects and computational systems. The negative consequences for society present a global challenge impacting every country and nearly all areas of public discourse. Disinformation bolsters authoritarians, weakens democratic voices and participation, targets women and marginalized groups, exploits and exacerbates existing social cleavages, and silences opposition. Across the networked public sphere¹ civil society, governments, and the private sector

are grappling with these new online threats and working through their own networks and with each other as part of a whole-of-society effort to improve the integrity of our information environment.

While disinformation has long been a challenge to democracy, the digital age necessitates a renewed commitment and fresh urgency to match the scale, speed, and pervasiveness of online information threats. Meaningful access to a healthy information environment is integral to the functioning of free, rights-respecting societies; as such, countering disinformation and promoting information integrity are necessary priorities for ensuring democracy can thrive globally in the next century and beyond.

HOW THE WORK WAS CONDUCTED

This guide is an ambitious effort to take a global look at measures to combat disinformation and promote information integrity—a collaborative examination of what is being done, what is working, and who is doing it. This resource has been developed by the International Foundation for Electoral Systems, the International Republican Institute, and the National Democratic Institute with support from [USAID \(https://www.usaid.gov/\)](https://www.usaid.gov/) to the [Consortium for Elections and Political Process Strengthening \(https://www.ndi.org/CEPPS\)](https://www.ndi.org/CEPPS), and is intended to serve as a guide for practitioners, civil society, and government stakeholders working to advance information integrity and strengthen societal resilience. The research has been conducted over two years, led by experts from all three organizations. Case studies have been conducted in three countries and the database includes more than 275 entries across over 80 countries in all regions outside of Antarctica, which will be updated and expanded over time. More than a dozen external experts have served as peer reviewers and editors. Due to COVID-19, some research efforts were curtailed. Due to the scope and scale of the challenge, the research is thorough, but not exhaustive. While drafting, actions by social media platforms, governments, civil society actors, and activists continued to evolve. As a result, we intend that this guide should be a living platform with substantive chapters updated annually and the Global Database updated more regularly.

WHAT'S IN THE GUIDE

Examples and quotes from all three cases are integrated throughout the guidebook to illustrate lessons learned and the evolution of counter-disinformation programming and other interventions. Where possible, links are provided to entries in the Global Database of Informational Interventions. The database is the most robust effort in the democracy community to catalogue funders, types of programs, organizations and descriptions of the project. In addition, topics include quotes

HIGHLIGHT

Over two years, CEPPS conducted in-country research in Colombia, Indonesia, and Ukraine. Key actors were interviewed

from interviews of stakeholders, reviews and analyses of programs, and reports on monitoring and evaluation. Media reports and academic literature focused on impact and effectiveness have also been included. Finally, this guide is intended to be a living document, and the database and the topics will be periodically revised and improved to reflect the ongoing evolution of the online and real world environment.

The topics are divided into three broad categories, examining the **roles** of specific stakeholder groups, legal, normative and research **responses**, as well as a crosscutting issues for addressing disinformation targeting women and marginalized groups, and elections. The topics include:

The topics include:

ROLES

- **Building Civil Society Capacity to Mitigate and Counter Disinformation** (/node/2690) looks at various efforts by civil society organizations to combat disinformation and promote information integrity through programs and other initiatives including fact checking, media literacy, online research and a host of other methods.
- **Helping Political Parties Protect the Integrity of Political Information** (/node/42/) explores the impact of disinformation and hate speech campaigns on political parties in developing countries and provides policy recommendations for parties in countering harmful forms of content and promoting positive ones.
- **Platform Specific Engagement for Information Integrity** (/node/2722/) explores varying policy, enforcement, and partnership responses by social media platforms (large and small) to address disinformation challenges.
- **Election Management Body (EMB) Approaches to Countering Disinformation** (/node/31/) explores the varying roles that EMBs play in countering disinformation and

based on their experience developing interventions, their role in the political system as well as their perspective on the information landscape and other related issues. These countries were chosen based on the relevant interventions and programs they have developed, demographic and geographic diversity, risk of foreign intervention, as well as critical recent elections and other political events. Ukraine is on the front lines of information space issues, as a civil war triggered by the Russian invasion of Crimea has created a contested information space, often influenced by the Kremlin. A massive, important Asian democracy, Indonesia has had recent elections in which social media played a critical role, spurring innovative responses from election management bodies and civil society to mitigate the impacts of disinformation and promote a healthy information environment. Colombia represents the final example, a country with both recent elections and a major peace agreement between the government and rebel groups ending a decades-long war. The pact's negotiation, a failed referendum, and finally ratification by the legislature have followed in successive years and provides an important case study of how a peace process and reconciliation are reflected and negotiated online alongside elections and other political events.

offers proactive, reactive and collaborative strategies for election authorities to consider.

- **Exposing Disinformation through Election Monitoring (/node/2716/)** examines the work and methods of international and domestic monitoring of the information space as a component of election observation.

RESPONSES

- **Developing Norms and Standards on Disinformation and Information Integrity Issues (/node/2743/)** provides an overview of global norms and standards that have been developed to counter disinformation that are consistent with human rights.
- **Laws, Regulations, and Enforcement Mechanisms (/node/2704/)** explores ways that national legal frameworks governing elections address social media, and provides a resource for lawmakers and international donors considering alterations to their own electoral frameworks.
- **Research and Evaluation Tools (/node/2749/)** for countering disinformation explores a variety of research tools that practitioners use to understand threat actors, targets, the information ecosystem, and program impact.

CROSSCUTTING DIMENSIONS

- **Understanding the Gender Dimensions of Disinformation (/node/13/)** explores how disinformation campaigns, viral misinformation and hate speech target and particularly affect women and people with diverse sexual orientations and gender identities by exploiting and manipulating their self-identities. As a result, this section and every other topic include considerations for gender and marginalized groups in programming and other interventions.

The [Database of Informational Interventions \(/interventions\)](/interventions) provides a comprehensive set of interventions that practitioners, donors, and analysts can use globally in understanding and countering disinformation.

9 BIG TAKEAWAYS

In conducting this analysis and looking at these critical aspects of the problems, the research team has identified key takeaways that should drive disinformation efforts going forward.



1 Disinformation exists in every information ecosystem in the world. No actor can address this alone. For this reason, a whole-of-society approach is needed that encourages actors from governments, civil society, and industry to work together to counter disinformation and strengthen societal response.



2 Countering disinformation is not THE top priority for most institutions, governments, political parties, or civil society groups. However, some of these actors proliferate both disinformation and misinformation. Until this sense of urgency drives a collective effort to address it, lasting change cannot be achieved.



3 Efforts to combat disinformation in elections and to combat existing societal cleavages are distinct but overlapping challenges. Donors and implementers should not let a bias toward technologically innovative programming undercut continued investment in building the types of durable capacity that make democratic stakeholders more resilient when disinformation challenges arise.



4 Public and private institutions such as Election Management Bodies (</topics/embs/0-overview-emb-approaches>) and platforms (</topics/platforms/0-overview-platforms>) are often well equipped to address disinformation challenges but lack credibility. By contrast, civil society (</topics/csos/0-introduction-building-civil-society-capacity>) is a credible actor, nimble, and essential but chronically under resourced.



5 No one approach (media literacy, fact checking, research and monitoring ([/topics/surveys/0-executive-summary/](#)), social media take downs, etc) is sufficient. A holistic approach to countering disinformation is essential.



6 Focusing on major events, such as the outcomes of elections and referendums ([/topics/monitoring/0-overview-election-monitoring/](#)), are effective in creating safe political processes. This contributes to, but does not achieve, a healthy information ecosystem.



7 Understanding the impact of gendered disinformation ([/topics/gender/0-overview-gender-disinformation/](#)) and the role gender plays in information integrity is critical. As such, interventions must include gender component and be localized for greater context from program design to implementation in order to increase effectiveness and minimize potential harm.



8 Disinformation efforts that rely on content moderation structures alone are not sufficient. Development of norms and standards ([/node/2743/](#)), legal and regulatory frameworks ([/node/2704/](#)) and better content moderation of social media platforms ([/node/2722/](#)) must be addressed in order to create a healthy information ecosystem. This is especially important to

strengthening complex information environments in the Global South.



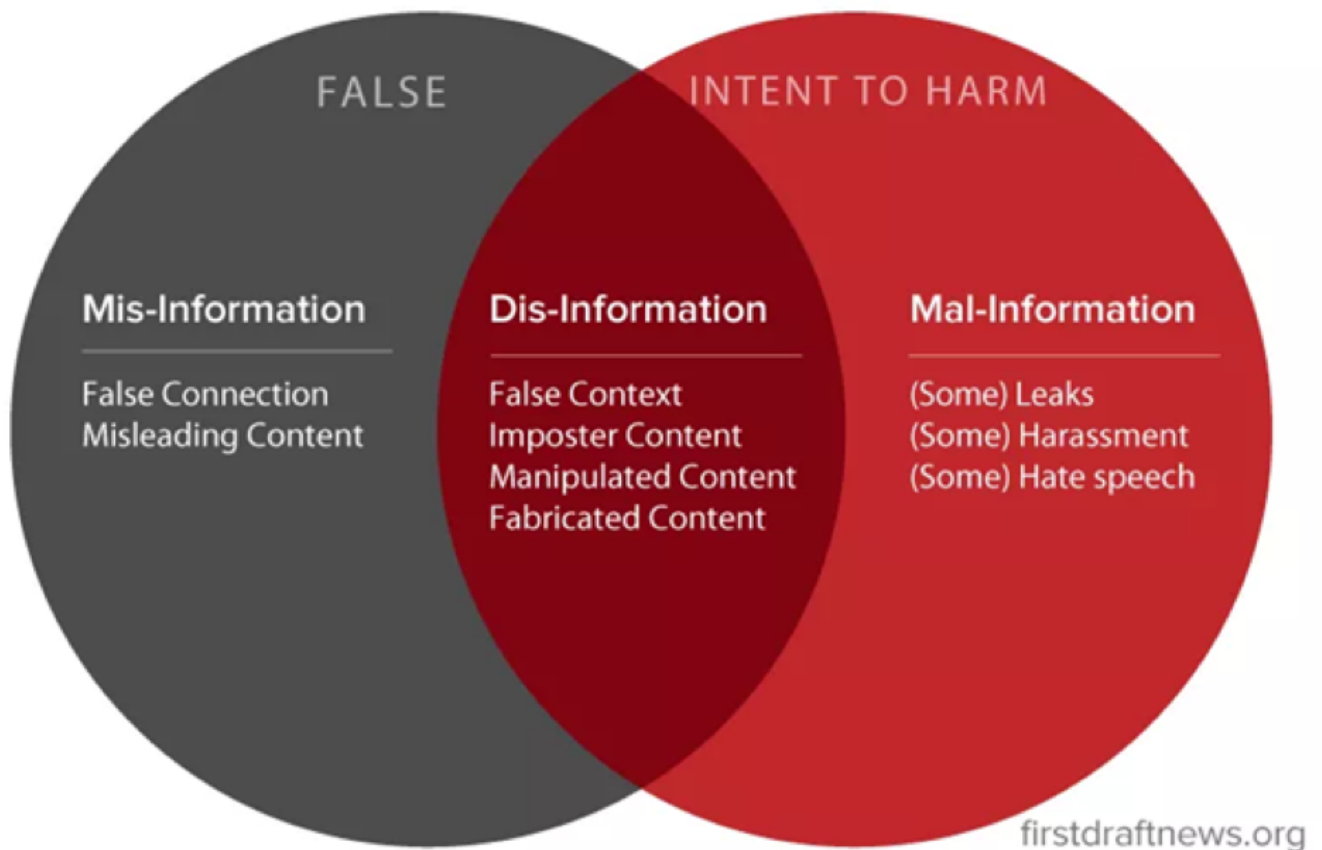
9 Parties (/node/42/) play a critical role in both political systems and the creation and dissemination of online campaigns that often propagate disinformation and other harmful forms of content. It is important that frameworks be put in place that discourage political parties from engaging in disinformation.

UNDERSTANDING DISINFORMATION

Significant work has been done in recent years to conceptually understand and diagnose information disorder. To conceptually ground our analysis, this guide builds its definitions and understanding of the problems and as well as solutions primarily on the work of [Data and Society](https://datasociety.net/) (<https://datasociety.net/>), [First Draft](https://firstdraftnews.org/) (<https://firstdraftnews.org/>), and the [Oxford Internet Institute's Computational Propaganda Project](https://comprop.oii.ox.ac.uk/) (<https://comprop.oii.ox.ac.uk/>). These three foundational resources are well-regarded in the broader community analyzing disinformation, as well as for the ways in which their conceptual frames lend themselves to adaptation for practical application.

First Draft's [Information Disorder](https://www.coe.int/en/web/freedom-expression/information-disorder) (<https://www.coe.int/en/web/freedom-expression/information-disorder>) provides clear definitions of information disorder, implications for democracy, the role of television, implications for local media, microtargeting, computational amplification, filter bubbles and echo chambers, and declining trust in the media and public institutions. The framework also describes how misinformation (information passed without the intent to deceive), disinformation (incorrect information passed with intent) and malinformation (true information made public with the intent to harm) are all playing roles in contributing to the disorder, which can also be understood as contributing to the corruption of information integrity in political systems and discourse.

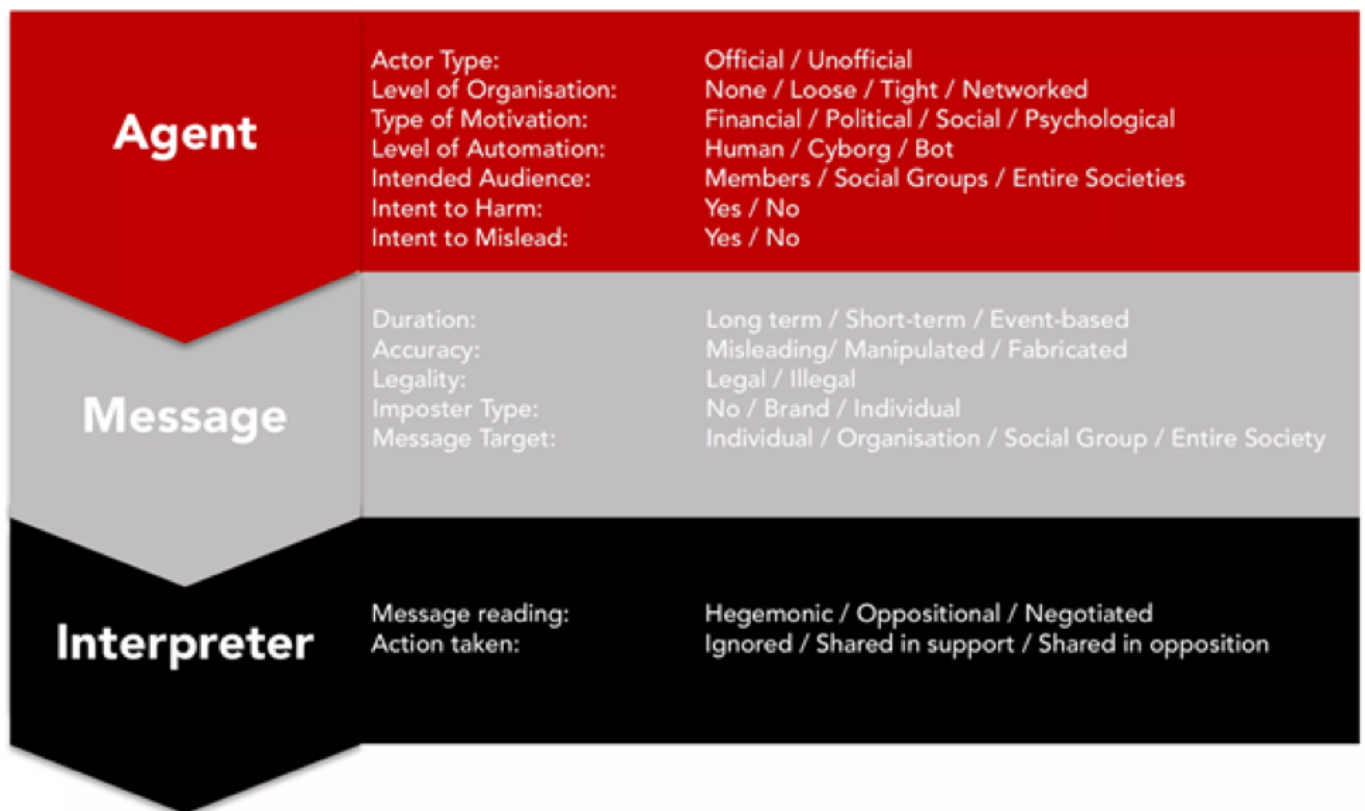
INFORMATION DISORDER



From Wardle, Claire, and Hossein Derakhshan. "Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking." Council of Europe, October 31, 2017.

<https://shorensteincenter.org/information-disorder-framework-for-research-and-policymaking/>.

The Information Disorder framework also focuses on elements of the information ecosystem including the information agent (or producer), the message and the interpreter. Messages pass through several phases, namely creation, production, and distribution. These aspects allow for us to interpret different kinds of efforts, whether they focus on one element of these three components, some or even all of them. Legal and Regulatory frameworks and norms and standards can target all of these aspects, and different actors such as platforms, civil society organizations, and governments can design responses that address them in different ways. For instance, media literacy efforts target the interpreters, while content moderation focuses on the messages and agents.



From Wardle, Claire, and Hossein Derakhshan. "Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking." Council of Europe, October 31, 2017.

<https://shorensteincenter.org/information-disorder-framework-for-research-and-policymaking/>.

The Oxford Internet Institute (OII) developed the term "computational propaganda" and defines this practice as "the assemblage of social media platforms, autonomous agents, and big data tasked with the manipulation of public opinion."² This framework allows us to expand our understanding of threats in the online space beyond disinformation to other forms of manipulation online, whether automated or human. It also helps to frame the problem as one including technical, sociological, and political responses. To help understand the virality of disinformation, OII's work demonstrates how communications, behavioral, and psychological studies--as well as computer, data and information science--all play a role.

Data & Society (<https://datasociety.net/>)'s Oxygen of Amplification demonstrates how the traditional media play a role in amplifying false narratives, and how it can be manipulated to promote disinformation and misinformation in different ways. Another research group that brings together diverse aspects of media and data analysis as well as social science research also helps define terms and standards. Our glossary relies on Data & Society's report on the Lexicon of Lies (https://datasociety.net/pubs/oh/DataAndSociety_LexiconofLies.pdf)³, as well as the First Draft's Essential Glossary (<https://medium.com/1st-draft/information-disorder-part-1-the-essential-glossary-19953c544fe3>) from its study Information Disorder, and other sources that are cultivated through our global database of approaches (/interventions) and other literature, including guidance by USAID⁴ and other organizations including CEPPS. Technical, media, and communications concepts will be included in the sections and these key terms help describe the problem in a shared way.

Footnotes

¹ Benkler, Yochai. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press, 2006.

² Howard, P. N., and Sam Woolley. "Political Communication, Computational Propaganda, and Autonomous Agents." Edited by Samuel Woolley and Philip N. Howard. *International Journal of Communication* 10, no. Special Issue (2016): 20.

³ Jack, Caroline. "Lexicon of Lies: Terms for Problematic Information." *Data & Society*, August 9, 2017. <https://datasociety.net/output/lexicon-of-lies/> (<https://datasociety.net/output/lexicon-of-lies/>).

⁴ *Disinformation Primer*, Center for Excellence on Democracy, Human Rights and Governance, USAID, February 2021.

BUILDING CIVIL SOCIETY CAPACITY TO MITIGATE AND COUNTER DISINFORMATION

0. OVERVIEW - CIVIL SOCIETY (/TOPICS/CSOS/0-INTRODUCTION- BUILDING-CIVIL-SOCIETY-CAPACITY)

Written by Amy Studdart, Senior Advisor for Digital Democracy at the International Republican Institute

Civil society approaches to countering disinformation encompass a variety of program types, including fact-checking, digital forensics and research, advocacy to governments and platforms, digital and media literacy, networking and coalition building, and international cooperation. Across these program approaches, implementation by civil society organizations (CSOs) has several advantages that could plausibly increase the effectiveness of programs. Civic groups can rapidly innovate, they are more closely connected to citizens that disinformation affects, better placed to understand its immediate impact, and able to build trust with local communities – a key factor in responding to specific information disorders – and more likely to be perceived by all parties as relatively objective. More specifically, civic associations promote the cooperation of citizens from distinct interest and identity groups, such as women, ethnic minorities, and persons with disabilities. As such, among key stakeholders, these organizations and coalitions are often best placed to identify disinformation campaigns that target marginalized groups or that exploit

existing gender norms or social divisions, and to mobilize broad opposition and responses to these campaigns. Across countries and global regions, civic groups have designed and implemented the following types of counter-disinformation programs:

Fact-checking (</topics/csos/2-fact-checking>)

Fact-checking initiatives attempt to identify and correct false or misleading information propagated either by political and economic elites or through peer-to-peer interactions on social media or messaging apps. Civic groups are uniquely placed to implement these programs for two related reasons: first, by acting as relatively objective, dispassionate sources, CSOs can be sources for corrections, especially given the highly politicized nature of disinformation campaigns. Second, CSOs tend to be less constrained, especially relative to journalists, in both methods and solutions.

Identifying Disinformation Narratives, Assets, and Coordinated Inauthentic Behavior (</topics/csos/3-identifying-disinformation-narratives-assets-and-coordinated-inauthentic-behavior>)

CSOs, often in collaboration with academics or research organizations, have played a prominent role in uncovering information operations. Civic groups have identified ongoing information operations around elections, identified coordinated inauthentic behavior for platforms, and conducted media monitoring to identify key information narratives. CSOs are often particularly well-placed to support the uptake and utilization of outputs from sophisticated research approaches, ensuring that findings are quickly actionable for decisionmakers or targets of disinformation campaigns. Furthermore, as women and other marginalized groups are often early targets of emerging campaigns, civic groups that represent these interests are often best placed to identify the emergence of these tactics, and to advocate for effective responses.

Advocacy Toward Platforms (</topics/csos/4-advocacy-toward-platforms>)

In their role as a mediator between citizens and governments, CSOs have a natural function of advocacy. Specifically, CSOs are well placed to identify how disinformation campaigns target and harm marginalized groups, which might not otherwise be obvious to the platforms themselves, and subsequently to advocate for platform policy changes that respond to those specific issues. However, civic groups face several challenges in advocacy toward media outlets and digital platforms, including strong platform financial incentives, limited access to decisionmakers, and knowledge gaps within civic groups. Network and coalition-based approaches to advocacy, particularly internationally, can help overcome these challenges by increasing leverage through collective action, including by amplifying the voices of marginalized groups and linking their priorities to broader policy goals

Advocacy Toward Governments (</topics/csos/5-advocacy-toward-governments>)

Civil society plays two critical roles vis-à-vis government responses to disinformation: (1) advocating for pro-democratic policies that protect and advance information integrity, including the equal value and equal rights of association for marginalized groups whose participation perpetrators of disinformation often seek to undermine, and (2) ensuring that responses to disinformation, information operations, and other information disorders do not clamp down on

free speech, access to information, or participatory politics in ways that might harm democratic processes and principles, again with a focus on how restrictions on association and expression often disproportionately affect marginalized groups. Again, the perception of CSOs as relatively objective can increase their credibility with decisionmakers, and collective action between organizations can make advocacy campaigns more effective.

Public Awareness/Media Literacy Campaigns (</topics/csos/6-public-awarenessmedia-literacy-campaigns>)

CSOs' connection to local communities and position as a relatively trusted source of information make them ideally placed to design and implement public awareness and media literacy programs. These interventions are implemented under the assumption that if audiences can utilize necessary critical thinking skills while consuming online and traditional media content, it will increase their ability to differentiate between factual and misleading or fake content. While the internet and social media platforms have improved access to media and information, as well as the plurality of news sources, they have nonetheless contributed to a decline in the quality of news and information. Improved media and digital literacy among audiences could play a significant role in helping reduce susceptibilities to disinformation overtime. Public awareness campaigns by civic groups can also help create perceptions of shared interests, particularly where they highlight how disinformation campaigns affect the democratic rights or engagement of women and other marginalized groups that might not otherwise be visible.

Building Trusted Networks for Accurate Information (</topics/csos/7-building-trusted-networks-accurate-information>)

CSOs have been critical in serving as a trusted source of information, particularly in environments in which state media or the government are the main perpetrators of disinformation, and in which the active propagation of disinformation is accompanied by censorship. While "word of mouth" and other creative information distribution activities have always been present in closed societies, those channels have taken on greater formality and scale as digital technologies, and particularly encrypted group chat applications, have become widely accessible.

International Collaboration (</topics/csos/9-international-collaboration>)



HIGHLIGHT

However, international collaboration, especially in terms of philanthropy and development assistance, should consider limitations imposed by small grants and short timelines. Responding to information disorders, or building resilience to them in the first place, may require infrastructure with high startup costs, and long-term ongoing support to ensure these initiatives are sustainable.

International cooperation is a critical factor behind civil society success. In addition to the leverage issue vis-à-vis companies discussed in this chapter, international cooperation allows civil society to share best practices in the rapidly evolving fields of digital forensics and counter-messaging, and to share information about emerging transnational threats and the proliferation of disinformation toolkits used by malign actors, both foreign and domestic.

Programmatic Recommendations (</topics/csos/10-conclusion-and-recommendations>)

Civic organizations play a key role in identifying and responding to information disorders, especially where they can establish reputations as relatively independent, objective actors. However, these advantages come with tradeoffs, especially if their constituencies tend to be relatively urban, highly educated, wealthier, or more internet-connected on average. Program designs should take care to target interventions to encourage uptake among underserved groups.

Network and coalition approaches to countering disinformation, including international collaboration, can identify comparative advantages, increase scale, and improve the diversity of programmatic approaches.

Relatedly, programs focused on civil society should incorporate an intentional focus on inclusion, and more specifically, the intersectionality of multiple marginalized identities, particularly in coalition and network approaches. Support for civic groups should incorporate a distinct analysis to identify unique challenges faced by individuals facing multiple forms of marginalization within a specific historical context, since perpetrators of disinformation campaigns may rely on the apathy or complicity of non-marginalized identity groups. Collective action is more likely when these groups and individuals that are not politically or socially marginalized understand that they have an interest in defending the rights of minority and marginalized groups.

Civic organizations may consider partnering with existing political or social institutions to scale programmatic responses to disinformation, especially if the organization itself has a small or narrow audience. One example might include partnering with school systems to implement media-literacy programs.

Programs working on advocacy, especially around internet or platform regulation, should consider the specific cultural context of debates surrounding tradeoffs between free expression and security.

Programs working with civic organizations to implement counter-disinformation programs should consider dedicated security training components, including cybersecurity, data protection, response plans for information attacks, and physical security from retaliation.

BUILDING CIVIL SOCIETY CAPACITY TO MITIGATE AND

COUNTER DISINFORMATION

1. INTRODUCTION (/TOPICS/CSOS/1-INTRODUCTION)

The role of civil society in fighting disinformation is multifaceted: fact-checking, digital forensics and research, advocacy to governments, advocacy to platforms, digital literacy campaigns, reconciliation, and international cooperation.

While definitions of civil society vary widely, and indeed there is significant debate about what does and does not constitute civil society, Larry Diamond, a senior fellow at the Freeman Spogli Institute for International Studies at Stanford University, provides a conceptualization that corresponds closely to democracy, rights, and governance (DRG) practitioners understand the concept:

“Civil society is...the realm of organized social life that is voluntary, self-generating, (largely) self-supporting, autonomous from the state, and bound by a legal order or set of shared rules. It is distinct from “society” in general in that it involves citizens acting collectively in a public sphere to express their interests, passions, and ideas, exchange information, achieve mutual goals, make demands on the state, and hold state officials accountable. Civil society is an intermediary entity, standing between the private sphere and the state. Thus, it excludes individual and family life, inward-looking group activity (e.g., for recreation, entertainment, or spirituality), the profit-making enterprise of individual business firms, and political efforts to take control of the state.”¹

Pointedly, civil society (as an ideal type) creates what political scientists call “cross-cutting cleavages” – overlapping identities that transcend narrow identities or interest groups based on gender, economic class, race or ethnicity, religious affiliation, sexual orientation, or political affiliation². Association through civic groups creates familiarity and a sense of shared interests between members of disparate and narrow identity groups. With regard to responding to disinformation, relative to these other forms of social organization that Diamond identified, civil society actors benefit from a number of advantages: they are more able to rapidly innovate than governments, technology companies, or media organizations; they are closer to those most impacted by disinformation, more likely to understand its immediate impact, and better able to build trust with impacted communities; their grassroots, localized knowledge is critical to rebuking false narratives; and, unlike governments or political actors, many civil society groups are less likely to be perceived as having a vested interest in propagating or counteracting political disinformation. One important potential strength of civic organizations for responding to disinformation is their capacity to generate shared interests and goals between disparate identity groups. As disinformation often disproportionately (and often earlier) targets women and

historically marginalized groups within specific contexts, CSOs or coalitions are often best placed to identify emerging campaigns early, and to generate awareness, mobilize opposition, or advocate responses broadly. By creating this sense of solidarity and shared interest, civic organizations are well placed not only to defend vulnerable groups from specific harms, but to increase the resilience to disinformation of society broadly, including members of groups who have not been historically vulnerable or marginalized. For all these reasons, civil society plays a critical role in the broader ecosystem for countering disinformation.

This chapter runs through a number of those interventions, details civil society's advantages and disadvantages as it relates to each intervention and concludes with recommendations – many of which are pulled from those indicated throughout the chapter – as to how to support and strengthen civil society's contributions to addressing disinformation.

BUILDING CIVIL SOCIETY CAPACITY TO MITIGATE AND COUNTER DISINFORMATION

2. FACT-CHECKING (/TOPICS/CSOS/2- FACT-CHECKING)

Many of the most successful and reliable fact-checking initiatives have been driven and staffed by independent media or trained journalists. Those actors are best placed to understand how to thoroughly investigate misleading content, reliable sourcing, and communicating in a dispassionate way about how and why a piece of content or a particular narrative is misleading. However, this is also a space in which civil society organizations have played a critical role.

First, CSOs often complement fact-checking initiatives (<https://www.dw.com/en/civil-society-actors-offer-community-based-fact-checking/a-53956502>) by acting as sources of information. Where journalists do not have firsthand knowledge of an issue, community, or geographical area subject to disinformation, civil society plays an essential role in either helping journalists debunk a claim through sharing their expertise, or in identifying the ways in which disinformation is impacting, for instance, marginalized communities (<https://medium.com/political-pandemonium-2020/how-civil-society-can-combat-misinformation-and-hate-speech-without-making-it-worse-887a16b8b9b6>). Given that disinformation disproportionately targets wedge issues in society, this second role is particularly important.

Secondly, because civil society is less constrained than journalists in terms of methodology and available solutions, they have a wider surface area on which to innovate. For instance, the spread of disinformation on



HIGHLIGHT

encrypted private messaging apps was an issue that caused so much consternation that many argued for the end of encryption altogether

In India, after hate speech and disinformation on WhatsApp led to real-world violence and loss of life, Facebook – WhatsApp’s parent company – limited group sizes and message forwarding. Multiple governments have shut down encrypted messaging platforms at various points. And even advanced democracies have started to demand – and even legislate – to create encryption backdoors for law enforcement.

(<http://cyberlaw.stanford.edu/blog/2020/01/earn-it-act-how-ban-end-end-encryption-without-actually-banning-it>).

FEATURED INTERVENTION COFACTS (/INTERVENTIONS/COFACTS)

A project of the g0v civic technology community in Taiwan, CoFacts is a fact checking bot for messaging groups. Messages can be forwarded to the CoFacts bot for fact checking by a team of volunteers; the CoFacts bot can also be added to private group

Similarly, in Ukraine, civic groups have led the development of fact-checking initiatives to counter both Russian propaganda and domestic disinformation.

FEATURED INTERVENTION STOPFAKE (/INTERVENTIONS/STOPFAKE)

The flagship project of their organization “StopFake” is currently well known to media professionals all over the world. Not only does it identify cases of fake information about events in Ukraine, but also actually initiated an international

FEATURED INTERVENTION

VOXCHECK

(/INTERVENTIONS/VOXCHECK)

Provides fact checking, explainers, and analytical articles, especially on issues of economic reform in Ukraine.

VoxUkraine is a non-profit digital media platform with a focus on economic issues. As part of its services, which also include research, analytical reports, explainer journalism, and economic education initiatives, its VoxCheck service uses a staff of experts to verify politicians' public statements on economic issues. The non-profit, civic orientation of these outlets provides several advantages; these fact-checking initiatives are situated within larger initiatives that focus on advocacy, journalism, public education, and media literacy. Furthermore, as digital outlets, they are largely able to retain more editorial independence than television, radio, and print outlets. However, these advantages entail tradeoffs. Representatives of VoxCheck, for example, noted that while they had a large audience, it was situated primarily in the capital of Kyiv, and was composed of younger, wealthier, and more educated consumers, who may already be likely to agree with their reports.⁴

Hundreds of civil society fact-checking initiatives have sprung up over the last five years around specific flashpoints, with the lessons learned and infrastructure built around those flashpoints then being applied to other issues that impact the same information ecosystem. Among the most systematic forums of international collaboration is the [International Fact-Checking Network \(IFCN\)](https://www.poynter.org/ifcn/) (<https://www.poynter.org/ifcn/>)⁵, a program at the Poynter institute that brings together factcheckers, provides training, creates basic standards for fact-checking, and advocates for factcheckers worldwide. The group also facilitates informal, reactive collaboration: in

May 2020, a group in France shared a story with the IFCN that alleged that the Italians had found a



DESIGN TIP

Civic groups considering fact-checking initiatives should consider being intentional about identifying new audiences, particularly those that might not be otherwise inclined to engage social media.

way to potentially cure COVID-19. Within an hour, other groups across Europe shared evidence of the same false story circulating in other countries, and their own evidence debunking the story. (<https://www.the-american-interest.com/2020/06/02/activists-against-digital-lies/>).

FEATURED INTERVENTION IFCN (/INTERVENTIONS/INTERNATIONAL- FACT-CHECKING-NETWORK)

The International Fact-Checking Network is a unit of the Poynter Institute dedicated to bringing together fact-checkers worldwide. The IFCN was launched in September 2015 to support a booming crop of fact-checking initiatives by promoting best

FEATURED INTERVENTION ANIMAL POLÍTICO (/INTERVENTIONS/ANIMAL- POLITICO)

Animal Político is a digital native medium that brings together journalists, designers, programmers and video editors to create content with rigor, precision and thought to serve citizens.

During the 2018 Mexican general elections, a CSO-driven initiative, Verificado 2018 (<https://verificado.mx/que-es-verificado-2018/>), partnered with Pop-Up News (<https://popup.news/>), Animal Político (<https://www.animalpolitico.com/>), and AJ+ Español (<https://www.facebook.com/ajplusespanol/>), along with 80 other partners (<https://verificado.mx/hasta-luego-hoy-cierra-verificado-2018/>) to fact-check and distribute election-related information, particularly among youth. Before the elections, Verificado was established as a youth civil society group, Verificado19S (<https://verificado19s.org/>), named in reference to the September 19, 2017 Puebla earthquake (<https://apnews.com/70b3a90e267d44138eb30203d96aab7d>) that caused much destruction in the Mexican states of Puebla and Morelos and the Greater Mexico City area, leading to hundreds of

deaths. The fact-checking initiative reached more than 200,000 followers on Facebook and Twitter and over 10,000 WhatsApp subscribers. Verificado19S aimed to gather and provide information regarding the earthquake from eyewitnesses through an [online questionnaire \(https://docs.google.com/forms/d/e/1FAIpQLSeu6rm6OocQiL0H73kw2mH62R3vgGjjpr6cAw3w3j-vhtEYcw/viewform\)](https://docs.google.com/forms/d/e/1FAIpQLSeu6rm6OocQiL0H73kw2mH62R3vgGjjpr6cAw3w3j-vhtEYcw/viewform). Verificado 2018 then utilized the infrastructure and reputation built around the earthquake to replicate a similar initiative around the elections. The initiatives filled an information vacuum in the absence of government-led initiatives and other trusted, reliable sources of information. The initiative received a broad base of financial support from Facebook, Google News Initiative, Twitter, Open Society Foundation, Oxfam México, and Mexicanos contra la Corrupción y la Impunida, further expanding its reach and ensuring the real and perceived independence of the initiative.

Colombia has similarly developed strong fact-checking and research groups focused on the online space that integrate fact-checking. A network of journalists known as the "[Editorial Board \(https://consejoderedaccion.org/\)](https://consejoderedaccion.org/)" (*Concejo de Redacción*) supports various journalistic initiatives including training and investigation support as well as fact-checking, and supports a group called ColombiaCheck that works to fact-check political statements. This work is inspired partly by the model of [Chequeado \(https://chequeado.com/\)](https://chequeado.com/), a group based in Argentina. [ColombiaCheck \(https://colombiacheck.com/sobre-nosotros\)](https://colombiacheck.com/sobre-nosotros) began fact-checking information around the peace process negotiations between the government and the FARC rebel group in 2015, and has since continued to develop its methodology through subsequent elections and continuing political events⁶. [ColombiaCheck is certified by the Poynter Institute's International Fact-Checking Network and has worked to check content on Facebook as a third party fact-checker \(/topics/platforms/0-overview-platforms\)](#).

FEATURED INTERVENTION COLOMBIA CHECK (/INTERVENTIONS/COLOMBIA-CHECK)

Colombiacheck is a project of the Editorial Board , a non-profit, non-partisan organization that brings together more than 100 associated journalists in Colombia to promote investigative journalism. The project consists of a digital, open and

Latin America as a whole has developed strong fact-checking initiatives, including in Brazil where Agência Lupa represents one of the first initiatives that began in 2015 and is now integrating with the Folha de São Paulo's UOL network, the second largest online media network in the country. In the 2018 national elections, various organizations including Agência Lupa, Aos Fatos and

traditional media organizations worked to collaborate through [Comprova](https://firstdraftnews.org/) (<https://firstdraftnews.org/>), a joint initiative supported by First Draft, which is a global project to combat mis- and disinformation that also provides the information disorder framework this guide is partly based on. This is based on the "CrossCheck" model where various media organizations "cross-check" facts and confirm them jointly across platforms, [which has been replicated in France, Germany, Nigeria, Spain, the UK and the U](#) (<https://firstdraftnews.org/>). There is no shortage of successful fact-checking initiatives around the world, ranging from [Africa Check](#) (<https://africacheck.org/>), the [Cyber News Verification Lab](#) (<https://tl.hku.hk/staff/teaching-development-grants/tdg-627/>) in Hong Kong, [BOOM](#) (<https://www.boomlive.in/>) in India, [Checazap](#) (<https://enoisconteudo.com.br/checazap/>) in Brazil, the Centre for Democracy and Development [Fact Check archive](#) (<https://www.cddwestafrica.org/category/fact-check/>) in West Africa, and Meedan's [Check](#) (<https://meedan.com/check>) initiative in Ukraine. As part of CEPPS, Internews has supported various initiatives globally ranging from [Ethiopia](#) (<https://internews.org/story/fighting-false-information-help-save-lives>) to the Philippines and Turkey.

FEATURED INTERVENTION AGÊNCIA LUPA (/INTERVENTIONS/AGENCIA-LUPA)

The Magnifier is the first news agency in Brazil to specialize in journalistic technique known worldwide as fact-checking and was founded on November 1, 2015. Its business plan began

BUILDING CIVIL SOCIETY CAPACITY TO MITIGATE AND COUNTER DISINFORMATION

3. IDENTIFYING DISINFORMATION NARRATIVES, ASSETS, AND COORDINATED INAUTHENTIC BEHAVIOR (/TOPICS/CSOS/3-IDENTIFYING-

DISINFORMATION-NARRATIVES-ASSETS-AND-COORDINATED-INAUTHENTIC-BEHAVIOR)

While much of the work of uncovering information operations has been done by academia and private threat intelligence companies, international civil society has played a prominent role in uncovering information operations. Again, because of its role facilitating cooperation between members of potentially disparate groups, CSOs are often best placed to identify emerging campaigns that target vulnerable groups that might not otherwise be visible, and to mobilize responses.

The DC-based Digital Forensics Lab (DFRLab), for instance, has identified a number of coordinated information operations, with many of those operations designed to discredit elections. Over a one-month period, DFRLab published work exposing various forms of information operations in Ukraine (<https://medium.com/dfrlab/internet-marketers-exploit-facebook-ads-to-rent-facebook-accounts-in-ukraine-eee156e230bf>), Georgia (<https://medium.com/dfrlab/georgian-far-right-and-pro-government-actors-collaborate-in-inauthentic-facebook-network-730b9593a729>), and Nigeria (<https://medium.com/dfrlab/nigerian-government-aligned-twitter-network-targets-endsars-protests-5bb01a96665c>). Past work on Brazil, Colombia, Mexico,

(<https://www.atlanticcouncil.org/wp-content/uploads/2019/09/Disinformation-in-Democracies.pdf>)

El Salvador, Ecuador, and Bolivia has advanced understanding of disinformation actors in Latin America. Those investigations are critical to informing election integrity work. Domestic groups also play a critical role. In Colombia, groups such as Silla Vacía, Linterna Verde and Liga Contra Silencio have worked to explore the online space in both open networks such as Facebook and Twitter and more closed ones such as WhatsApp (<https://linternaverde.co/wp-content/uploads/2019/12/informe-whatsapp-FINAL-ENG-1OCT.pdf>) during elections, the referendum on its peace process, and other political events. As a specific example of how civic groups can identify emerging harmful narratives and link them to the interests of citizens more broadly, Linterna Verde has focused on online discourse focusing on female candidates online with the Liberty of the Press Foundation (Fundación para la Libertad de



HIGHLIGHT

In Ukraine, groups like StopFake have developed methods for digital exposure, reporting, and the public awareness-raising of campaigns, while groups such as Texty have collaborated with NDI (<https://www.ndi.org/our-stories/ukraine-new-way-battle-disinformation-meets-success-wins-awards>) to develop maps of networks, content, and critical trends within that context.

Prensa or FLIP (<https://flip.org.co/>) and how disinformation about women spreads online in the context of the 2018 presidential election (<https://flip.org.co/images/Documentos/informe-poligrafogenero-22mayo.pdf>).

Again, this early warning and response is important not only for protecting vulnerable groups that are the targets of these emerging campaigns, but to mobilizes responses in a way that maintains the integrity of the broader information ecosystem, including for members of groups that are not necessarily marginalized.

Digital forensics efforts are also being conducted by grassroots civil society organizations, and there is evidence of impact. For instance, days before the 2019 election in Moldova, Facebook removed over 100 accounts and pages identified by the civil society group (<https://freedomhouse.org/article/together-we-are-stronger-social-media-companies-civil-society-and-fight-against>), Trolless, as engaging in inauthentic behavior. Internews has also developed methods to track rumors in contexts starting in Liberia in 2014, which it has built into a detailed methodology (<https://internews.org/resource/managing-misinformation-humanitarian-context>) that is part of its learning collection of resources for training on disinformation and other media issues. However, a great deal of work needs to be done to ensure that local civil society groups have access to digital forensics expertise and the media monitoring tools that help researchers identify issues. NDI has developed the guide to Data Analytics for Social Media Monitoring (<https://www.ndi.org/publications/data-analytics-social-media-monitoring>) and translated it into Arabic, Portuguese and Spanish, partly to address this gap in the



HIGHLIGHT

While the field is, by its nature, very accessible, many of the resources that digital forensics researchers rely on, including how-to guides for beginners, are often only available in English or a limited set of languages and are not widely known.



HIGHLIGHT

As a specific example of how civic groups can identify emerging harmful narratives and link them to the interests of citizens more broadly, Linterna Verde has focused on online discourse focusing on female candidates online with the Liberty of the Press Foundation (Fundación para la Libertad de Prensa or FLIP (<https://flip.org.co/>)) and how disinformation about women spreads online in the context of the 2018 presidential election (<https://flip.org.co/images/Documentos/informe-poligrafogenero-22mayo.pdf>).

research community. [More examples are available in the Intervention Database. \(/interventions\)](#)

BUILDING CIVIL SOCIETY CAPACITY TO MITIGATE AND COUNTER DISINFORMATION

4. ADVOCACY TOWARD PLATFORMS (/TOPICS/CSOS/4-ADVOCACY-TOWARD- PLATFORMS)

Civil society advocacy is critical to changing platform product, policy, and resource allocation. It is also absolutely essential for raising concerns with platforms in ways that force action. Again, as perpetrators of disinformation often target context-specific wedge issues, including social and political cleavages, organizations that represent the interests of historically marginalized groups may be best placed to identify emerging issues that might otherwise not be obvious to platforms or ostensible regulators, and to advocate for reform.

In the U.S., a successful civil society advocacy effort led Reddit banned 2000 subreddits (forums dedicated to particular communities or interest areas), including [r/The_Donald](#), [r/gendercritical](#), and [R/ChapoTrapHouse](#). The decisions marked a major shift in policy. Previously, Reddit had functioned as an essentially libertarian space, with the rules of what was and was not allowed in each subreddit were set by moderators and creators of each subreddit rather than the platform itself. This led to some rather bizarre, sometimes delightful outcomes: in [one popular subreddit \(https://www.reddit.com/r/CatsStandingUp/\)](#), the only acceptable posts are pictures of cats standing up, and the only acceptable title or comment is “Cat.” The theory was that if a user disliked the content or community of a particular subreddit, they should simply find or establish another subreddit that they did like. However, as Reddit evolved from a niche place for absurd humor and shared interests into a major social media platform, disinformation, hate speech, and the affordances around community-building started to lead to real-world harms: the generation and popularization of conspiracy theories which would then platform jump and become viral, the abuse of the platform by malign actors, and coordination on the platform that led to offline criminal activity. Given that Reddit’s entire product is founded on the basis of community self-moderation, the ban marked a significant divergence in approach. While it is possible that the platform may have decided to take the step anyway, it is notable that [Reddit’s decision to quarantine r/The_Donald \(https://www.theverge.com/2019/6/26/18759967/reddit-quarantines-the-donald-trump-subreddit-misbehavior-violence-police-oregon\)](#) came two days after the US civil society group, Media Matters, launched a campaign to draw attention to how members of the subreddit were supporting attacks on police officers and public officials in Oregon.

Despite these nascent steps in the right direction, civil society groups and organizations outside of the United States and, to a lesser extent, Europe, are disadvantaged in their capacity to conduct effective advocacy vis-à-vis the platforms. Most successful attempts to change platform behavior – as in Myanmar (<https://dangerousspeech.org/dear-mark-global-civil-society-demands-that-facebook-act-against-dangerous-speech/>), Kenya (<https://qz.com/africa/1044573/facebook-and-whatsapp-introduce-fake-news-tool-ahead-of-kenya-elections/>), or Taiwan



HIGHLIGHT

CIVIC GROUPS, EARLY WARNING AND PLATFORM ADVOCACY

Facebook has established structured pathways for advocacy and input from civil society through its Civic Integrity and Global Insights program (</topics/platforms/0-introduction-platforms>)⁷, an initiative designed to solicit actionable input from grassroots communities around the world. These inputs are inherently limited in scope and are unlikely to lead to a radical shift in approach, but it has created a mechanism through which civil society in select countries are able to work with an interdisciplinary team to either get out ahead of issues, or rapidly resolve evolving threats to information integrity. This program and example of a mechanism through which civic groups, especially those representing women or marginalized groups, can advocate for platform responses to emerging disinformation campaigns, both to protect members of the groups they represent, but also to develop broader resilience of the information ecosystem.

(<https://www.nytimes.com/2020/06/11/technology/twitter-chinese-misinformation.html>) – have been accompanied by pressure from the U.S. government, civil society, or media. There are certain limitations that grassroots CSOs outside of the US face:

- Financial incentive: the U.S. is, for most companies, the biggest market in terms of financial return (although not absolute users or growth). As such, advocacy efforts in the U.S. and the negative PR those efforts generate impact consumer behavior, which directly impacts a given company's bottom line.
- The specter of regulation: for U.S. platforms, regulation coming out of Washington is sufficiently concerning enough that companies will often try to get ahead of the issues that voters care about and are thus most likely to lead to the kinds of regulation that can be harmful to business interests or operations.
- Cultural affinity: U.S. platforms and their employees are more clearly aligned with U.S. civil society than they are with civil society groups globally, and so critiques will land with more felt emotional weight in a way that can impact employee morale, lead to internal uprisings, or even resonate more clearly with leadership in a way that balances other interests. For instance, hate speech directed at African Americans is a more easily understood harm to companies staffed by Americans than is hate speech directed at Dalit's in India. Debates around freedom of speech are rooted in a U.S. cultural context, while concerns that lead with a desire for social harmony may not resonate as easily.
- Access: in many countries, even those in which the majority of the population uses a platform, the companies have, at best, sales and policy staff on the ground. Policy staff's principal roles are as lobbyists: they are rewarded on the basis of their ability to shape the regulatory environment in a way that benefits the company. They are not hired or rewarded for their relationships with civil society, and often struggle to navigate the complex web of interests of a given technology platform. At best, these limited touch points result in inaction. Far worse are those instances in which the company policy team in-country has interests which actively run counter to or may endanger civil society groups (for instance, where a group is critical of the government). In the U.S., meanwhile, civil society has multiple touchpoints with company representation, across teams and levels of seniority. As such, civil society in smaller markets struggles to find the right point of leverage within a company, even where those companies have teams designed to cover the issue of concern.
- Knowledge gap: civil society groups, particularly those working on issues not directly related to digital issues or disinformation, often lack sufficient knowledge of how technology platforms operate, the tools and resources they have to address issues, or the tensions endemic in and potential negative externalities surrounding decisions about content moderation.

Efforts such as the [Design 4 Democracy \(D4D\) Coalition](https://d4dcoalition.org/) (<https://d4dcoalition.org/>), which includes the National Democratic Institute (NDI), the International Republican Institute (IRI), International Foundation for Electoral Systems (IFES), and International IDEA, as well as a number of grassroots NGOs and the KeepItOn Coalition run by AccessNow, have started to address the challenge of leverage vis-à-vis the companies. By creating trusted avenues through which grassroots CSOs can work with higher capacity INGOs on advocacy efforts, the communication gap should theoretically become an easier one to bridge. However, a great deal of work needs to be done to ensure that companies further develop and invest in the teams they need to ensure that policy and product are responsive to the hyper-local information disorders that lead to negative outcomes.

Earlier in 2019, international pressure from several stakeholders, including society advocacy efforts, encouraged Facebook to increase oversight on political advertising, especially ahead of crucial elections in India, Nigeria, Ukraine, and the European Union. These efforts have led Facebook to "[extend some of its political advertising rules and tools for curbing election interference to India, Nigeria, Ukraine, and the European Union before significant votes](https://www.reuters.com/article/us-facebook-election-exclusive/exclusive-facebook-brings-stricter-ads-rules-to-countries-with-big-2019-votes-idUSKCN1PA0BT)" (<https://www.reuters.com/article/us-facebook-election-exclusive/exclusive-facebook-brings-stricter-ads-rules-to-countries-with-big-2019-votes-idUSKCN1PA0BT>). The web-based initiative [Media Matters for Pakistan](http://mediamatterspakistan.org/) (<http://mediamatterspakistan.org/>) also highlights independent efforts to hold mainstream media accountable to higher standards of journalism. This watchdog youth group raises awareness about the ethical and ideological issues found in media content and advocates against increased restrictions by the Pakistani government against digital media and freedom of expression. Similarly, the [EU DisinfoLab](https://www.disinfo.eu/) (<https://www.disinfo.eu/>) provides research and analysis on disinformation campaigns in the region, on traditional and online media platforms, to ensure that their advocacy efforts are "grounded in sound analyses." The initiatives mentioned above coupled with government actors to lead positive reforms to increase transparency. For more on platform engagement, [see the guide section on the subject \(/topics/platforms/0-overview-platforms\)](/topics/platforms/0-overview-platforms), or continue reading the section on building civil society capacity to mitigate and counter disinformation.



HIGHLIGHT

RESEARCH FOCUS: THE REGULATION- FREE EXPRESSION DILEMMA

In the course of the research for this project, several respondents identified potential free speech tradeoffs from regulation of digital platforms as a key ongoing policy debate. In Ukraine, for example, armed conflict with Russian-backed separatists in the country's eastern regions of Donetsk and Luhansk has created an acute need to balance free expression and national security. The government of Ukraine has banned Russian social media platforms and domestic television stations accused of disseminating pro-Russian propaganda, [earning rebukes](https://www.aljazeera.com/news/2021/2/5/ukraine-president-bans-pro-russian-networks-risking-support) (<https://www.aljazeera.com/news/2021/2/5/ukraine-president-bans-pro-russian-networks-risking-support>) from international organizations and international nongovernmental organizations that advocate for free media. While there is no clear consensus on the issue of platform regulation, civic advocacy groups are important conduits for channeling arguments to decisionmakers.

BUILDING CIVIL SOCIETY CAPACITY TO MITIGATE AND COUNTER DISINFORMATION

5. ADVOCACY TOWARD GOVERNMENTS (/TOPICS/CSOS/5-ADVOCACY-TOWARD- GOVERNMENTS)

Civil society plays two critical roles vis-à-vis government responses to disinformation: (1) advocating for pro-democratic policies that protect and advance information integrity, especially the protection of free expression and free association for marginalized groups and (2) ensuring that responses to disinformation, information operations, and other information disorders do not clamp down on free speech, access to information, or participatory politics in ways that might harm democratic processes and principles, given that these responses themselves may ultimately be used disproportionately to undermine the democratic rights of marginalized groups.

Government responses (/topics/csos/5-advocacy-toward-governments#guide), can – in the worst instances – include social media or internet shutdowns, heavy-handed regulation of online speech, or criminalization of certain types of online activity, all of which can backfire by infringing on civil liberties or exacerbating political inequity. Civil society thus serves not only as a useful counteractive force to those potential outcomes, but also as a space in which policy, technical, or social interventions can be tested, socialized, and iterated (<https://www.the-american-interest.com/2020/06/02/activists-against-digital-lies/?utm->

[access=newsletter&utm_source=TAI+Today&utm_campaign=330a1ea732-EMAIL_CAMPAIGN_2019_07_26_05_56_COPY_01&utm_medium=email&utm_term=0_6322a81c35-330a1ea732-178771625&mc_cid=330a1ea732&mc_eid=d42d924bff](https://www.the-american-interest.com/2020/06/02/activists-against-digital-lies/?utm-access=newsletter&utm_source=TAI+Today&utm_campaign=330a1ea732-EMAIL_CAMPAIGN_2019_07_26_05_56_COPY_01&utm_medium=email&utm_term=0_6322a81c35-330a1ea732-178771625&mc_cid=330a1ea732&mc_eid=d42d924bff)) before being subject to scale. Civil society is also unburdened with another challenge that governments have: given the often political nature of disinformation, and its utilization by political actors, incumbent governments



HIGHLIGHT

The Poynter's Institute's guide to anti-misinformation actions around the world (<https://www.poynter.org/ifcn/anti-misinformation-actions/>) details a range of policy experts initiatives to address the growing threat of disinformation.

often lack the real and perceived neutrality to ensure that responses are seen as fair, rather than as an attempt to undermine an opposition that may well be the principal beneficiary of disinformation.

Saudi Arabia (<https://www.arabnews.com/node/1668686/saudi-arabia>) threatened citizens and residents spreading rumors and fake news with five years jail sentence and hefty fines (<https://saudigazette.com.sa/article/545523>) sending a strong signal following the brutal killing of Washington Post columnist Jamal Khashoggi in 2018 at the Saudi embassy in Istanbul. In the same year, Ugandan (<https://www.theguardian.com/global-development/2019/feb/27/millions-of-ugandans-quit-internet-after-introduction-of-social-media-tax-free-speech>) officials introduced a "social media tax" that requires users to pay 200 Ugandan shillings a day to access specific online and social media platforms to tackle online gossip (<https://www.theguardian.com/world/2018/jun/01/social-media-use-taxed-in-uganda-to-tackle-gossip>). In Belarus (<https://freedomhouse.org/country/belarus/freedom-world/2020>), the parliament passed a law allowing the persecution of citizens who spread fake news. Organizations like the Committee to Protect Journalists (CPJ) and its partners (<https://cpj.org/campaigns/free-the-press/2020/>) have been the forefront of advocacy and policy reform efforts to support freedom of speech and to counter censorship efforts in places like South Africa (<https://cpj.org/2020/03/south-africa-enacts-regulations-criminalizing-disi/>) and Bolivia (<https://cpj.org/2020/04/bolivia-enacts-decree-criminalizing-disinformation/>) where leaders use disinformation as an excuse to jail journalists amid fears over the COVID-19 pandemic.

BUILDING CIVIL SOCIETY CAPACITY TO MITIGATE AND COUNTER DISINFORMATION

6. PUBLIC AWARENESS/MEDIA LITERACY CAMPAIGNS (/TOPICS/CSOS/6-PUBLIC- AWARENESSMEDIA-LITERACY- CAMPAIGNS)

Digital and media literacy interventions are implemented under the assumption that if audiences can utilize necessary critical thinking skills while consuming online and traditional media content, it will increase their ability to differentiate between factual and misleading or fake content. CSOs are particularly well placed to implement these programs because of the role of civil society in creating cross-cutting cleavages and shared interests. Beyond potential improvements in citizen capacity to identify false news, these programs can help raise awareness of how disinformation

narratives disproportionately harm women and marginalized groups. Plausibly, this shared awareness could help civic groups build broader support for advocacy or responses, although the evidence for the effect of these programs on citizen attitudes toward marginalized groups is yet unclear. These types of interventions aim to help audiences exercise caution and avoid blind trust of media content and other information available on the internet. The interventions are deployed in response to audiences not only consuming disinformation but also assisting in spreading such content to a larger group of audiences without efforts to verify content accuracy. The increasing media shift into the digital environment has proved to be a double-edged sword. The internet and social media platforms have improved access to media and information, as well as the plurality of news sources, but have nonetheless contributed to a decline in the quality of news and information. Improved media and digital literacy among audiences could play a significant role in helping reduce susceptibilities to disinformation overtime.

As some implementers identified through their work, much of the digital and media literacy and associated critical thinking skills start can and should be taught from a young age, similar to other necessary education skills. International Research & Exchanges Board (IREX) 's Learn to Discern (L2D) (<https://www.irex.org/project/learn-discern-l2d-media-literacy-training>) is one of the most successful media literacy initiatives that builds upon the point mentioned earlier. IREX has developed a media literacy curriculum that is taught in classrooms, libraries, and community centers in Ukraine, reaching over 62,000 individuals of all ages ([https://www.irex.org/sites/default/files/IREX%20Learn%20to%20Discern%20Results%20Factsheet%](https://www.irex.org/sites/default/files/IREX%20Learn%20to%20Discern%20Results%20Factsheet%202019-2020.pdf) approach adopted by IREX aims to build communities' resilience to resist disinformation, propaganda, and hate speech that is widespread in traditional and online media in Ukraine. After gaining much traction and success in Ukraine, L2D has been implemented in Serbia, Tunisia, Jordan, Indonesia, and the United States. With an interactive curriculum that engages audiences on the topic through games and multimedia content, the L2D initiative was able to attract young adults and raise awareness among them on the impact of disinformation on the lives of average citizens.

FEATURED INTERVENTION

LEARN TO DISCERN (/INTERVENTIONS/LEARN- DISCERN-L2D-MEDIA-LITERACY- TRAINING)

IREX's Learn to Discern approach helps citizens recognize and resist disinformation, propaganda, and hate speech. Learn to Discern's unique methodology builds practical skills for citizens of all ages through interactive training,

A year and a half after the kick-off of the project in Ukraine, IREX conducted [an impact evaluation survey \(/topics/surveys/4-evaluative-research-counter-disinformation-programs#evaluation\)](https://www.irex.org/sites/default/files/node/resource/impact-study-media-literacy-ukraine.pdf) in 2017 (<https://www.irex.org/sites/default/files/node/resource/impact-study-media-literacy-ukraine.pdf>), which reflected that 28% of L2D beneficiaries are "more likely to demonstrate a sophisticated knowledge of the news media industry" and 25% are "more likely to self-report checking multiple news sources." After piloting L2D-enhanced curricula in 2018 for over 5,000 students in the 8th and 9th grades in 50 schools, IREX evaluated their beneficiaries through a [survey \(https://www.irex.org/sites/default/files/node/resource/evaluation-learn-to-discern-in-schools-ukraine.pdf\)](https://www.irex.org/sites/default/files/node/resource/evaluation-learn-to-discern-in-schools-ukraine.pdf) that demonstrated that L2D students performed better than peers in a controlled group when "identifying facts and opinions, false stories, hate speech, and demonstrated a deeper knowledge of the news media sector." Since then, IREX has expanded the curricula to over 650 schools across Ukraine and collaborate with the Ukrainian Ministry of Education and Science to incorporate the curricula into the education system in Ukraine. IREX has received support from the Canadian government, the U.S. Embassy in Ukraine, and the U.K. Government's Department for International Development, and has partnered with the local organizations [Academy of Ukrainian Press \(https://www.aup.com.ua/en/mainen/\)](https://www.aup.com.ua/en/mainen/) and [StopFake \(https://www.stopfake.org/en/main/\)](https://www.stopfake.org/en/main/) to implement the L2D program since 2015.

Due to the increased attention on pro-Russian propaganda and disinformation, Ukraine and neighboring countries in Eastern Europe have served as the testing laboratory for a large number of countering disinformation initiatives. However, media and digital literacy initiatives have not been limited to Europe or to addressing Russian propaganda, and have taken many forms elsewhere around the world. The growing use of information and technology tools across Africa has brought about initiatives such as the [African Centre for Media and Information Literacy \(AFRICMIL\) \(https://www.africmil.org/\)](https://www.africmil.org/) aiming to educate youth on the effective use of those tools. AFRICMIL kicked off the first [Africa Media Literacy Conference \(https://www.africmil.org/programmes-and-projects/media-information-literacy/africa-media-literacy-conference/\)](https://www.africmil.org/programmes-and-projects/media-information-literacy/africa-media-literacy-conference/) in 2008 to further promote that goal. With support from the United Nations Educational, Scientific, and Cultural Organization (UNESCO), AFRICMIL has worked with the Nigerian youth to enhance their understanding of the impact of media and information consumption to increase their media literacy. The conference launched the [MIL University Network of Nigeria \(MILUNN\) \(https://www.africmil.org/unescoyouthmil/report-of-workshop/\)](https://www.africmil.org/unescoyouthmil/report-of-workshop/) to engage youth in Nigeria to be more critically aware of the role of media and information in their communities and provide awareness on the topic. The contribution made by AFRICMIL to raising awareness among journalists on ICT tools and creating a dialogue between peers locally and regionally across the content has proved to be instrumental in ensuring the voices of young people are heard. Egyptian fact-checking organization [Matsda2sh \(https://www.facebook.com/matsda2sh/\)](https://www.facebook.com/matsda2sh/) ("do not believe") has reached over 500 thousand followers on Facebook with awareness videos and photos highlighting the dangers of disinformation to the society with infographics and debunking statements with facts, including statements made by Egyptian President Abdel Fattah Al-Sisi.

In Indonesia, the anti-hoax grassroots civil society organization Masyarakat Anti Fitnah Indonesia (MAFINDO (<https://www.mafindo.or.id/>)) has led a CekFacta (<http://..CekFakta.com>), a content verification initiative site that promotes digital literacy among the public. MAFINDO's Facebook page (<https://www.facebook.com/MafindoID/>) has over 34,000 likes on their Facebook page through which it raises awareness on hoaxes and the dangers they pose to the community. MAFINDO has also worked on mapping out a popular hoax in 2018 and 2019 to enhance audiences' understanding of the malicious content that infiltrates their societies the most. The group has posted videos on their page that aim to highlight the dangers of hoaxes and false information; two of the videos uploaded on Facebook have reached over 32,000. However, despite the relatively large number of page followers and the traction that some of the group's content gets from audiences, recent posts have not received more than an average of a few hundred views and minimal likes and interaction from viewers. Moreover, another Indonesian group, Turn Back Hoax (<https://turnbackhoax.id/>), has more than 200,000 likes and followers on their Facebook page (<https://www.facebook.com/TurnBackHoax/>) and receives regular engagement on posts from followers.

FEATURED INTERVENTION CEKFACTA - MAFINDO (/INTERVENTIONS/CEKFACTA- MAFINDO)

MAFINDO is an anti hoax CSO (civil society organization). We began as an online grassroots movement in 2015. Founded as an organization on 19th November 2016



DESIGN TIP

In order to effectively evaluate the integrity of information to understand the needs and tailor programmatic responses to specific contexts, digital and media

Open source global initiatives such as the [Mozilla Web Literacy Framework](https://foundation.mozilla.org/en/initiatives/web-literacy/)

literacy efforts should be coupled with the media monitoring and verification initiatives explored in the next section.

(<https://foundation.mozilla.org/en/initiatives/web-literacy/>) and the [Facebook Digital Literacy Library](https://www.facebook.com/safety/educators/) (<https://www.facebook.com/safety/educators/>), where users can access educational literacy materials that can be accessed at any time and anywhere, offer an opportunity for users to learn how to effectively navigate the virtual world. Interactive games such as the [Bad News DROG](https://www.aboutbadnews.com/) (<https://www.aboutbadnews.com/>) supported by the [Dutch Journalism Fund](https://www.svdj.nl/dutch-journalism-fund/) (<https://www.svdj.nl/dutch-journalism-fund/>) takes users on a journey where users are asked to prove their credibility. Such interactive software serves as an educational tool. It provides a more digestible context for the dangers of disinformation in the daily lives of citizens and to society in general. The [News Literacy Project](https://newslit.org/) (<https://newslit.org/>)'s [Checkology](https://newslit.org/educators/checkology/) (<https://newslit.org/educators/checkology/>) initiative is built to support both students and educators and serves as an educational tool to provide comprehensive understanding to consumers of information. The project claims to have achieved significant results in the virtual classrooms as "more than two-thirds of students were able to identify the standards of quality journalism after completing Checkology lessons."

Digital and media literacy programs significantly helped with understanding audiences' consumption and in framing audiences' needs in order to build their resilience to false information, primarily targeted disinformation that aims to create divisions between citizens.

BUILDING CIVIL SOCIETY CAPACITY TO MITIGATE AND COUNTER DISINFORMATION

7. BUILDING TRUSTED NETWORKS FOR ACCURATE INFORMATION (/TOPICS/CSOS/7-BUILDING-TRUSTED- NETWORKS-ACCURATE-INFORMATION)

In information environments in which state media or the government are the main perpetrators of disinformation, and in which the active propagation of disinformation is accompanied by censorship, civil society has been absolutely critical in developing trusted networks and environments through which information can be shared. While "word of mouth" and other

creative information distribution activities have always been present in closed societies, those channels have taken on greater formality and scale as digital technologies, and particularly encrypted group chat applications, have become widely accessible.

In Zimbabwe, where state media dominates the media space, digital media groups such as [263 Chat \(https://263chat.com/\)](https://263chat.com/), established in 2012, capitalized on the increased use of digital platforms in the country to amplify the voices of citizens, increase their access, and encourage a dialogue among them. The group understood early on that with [WhatsApp use representing almost half of all internet traffic in Zimbabwe \(https://www.niemanlab.org/2019/03/whatsapp-has-come-in-to-fill-the-void-in-zimbabwe-the-future-of-news-is-messaging/\)](https://www.niemanlab.org/2019/03/whatsapp-has-come-in-to-fill-the-void-in-zimbabwe-the-future-of-news-is-messaging/) they can utilize it to package news information in a more digestible way that addresses the spread of disinformation in the country. As a result, 263 Chat distributes their e-paper for free to [more than 35,000 subscribers \(https://blog.wan- ifra.org/2019/07/31/how-zimbabwes-263chat-distributes-news-on-whatsapp\)](https://blog.wan- ifra.org/2019/07/31/how-zimbabwes-263chat-distributes-news-on-whatsapp) on WhatsApp. The founder of 263 Chat, [Nigel Mugamu \(https://twitter.com/SirNige\)](https://twitter.com/SirNige), has more than 100 thousand followers on Twitter, and [263 Chat's Twitter account \(https://twitter.com/263chat\)](https://twitter.com/263chat) has close to half a million followers, an impressive number for a platform now widely used in Zimbabwe.

FEATURED INTERVENTION

263CHAT

(/INTERVENTIONS/263CHAT)

263Chat was launched on September 29 2012 as a way of encouraging and participating in progressive and national dialogue in Zimbabwe. The use of the internet and the numerous social media tools available play an integral role in this entire process.

A number of similar initiatives exist in Venezuela, a country in which the public information space is almost entirely dominated by government propaganda and censorship. A number of civil society groups and independent activists have created WhatsApp channels, sometimes consisting of several hundred members, through which verified, reliable, and trusted information is transmitted. Those channels have played an interesting role during the COVID-19 pandemic. While they were originally created to address specific issues of concern to given civil society groups, these networks have since been used as distribution channels for accurate health information, including statistics about the virus' spread, and public service announcement advice about how to avoid contracting the virus.

BUILDING CIVIL SOCIETY CAPACITY TO MITIGATE AND COUNTER DISINFORMATION

8. CIVIL SOCIETY AS TARGETS OF DISINFORMATION (/TOPICS/CSOS/8- CIVIL-SOCIETY-TARGETS- DISINFORMATION)

Most of this chapter has explored civil society interventions that can address challenges to information integrity. Another important consideration, however, is how civil society organizations, their beneficiaries, and the issues they work on often become the targets of disinformation campaigns.

This has a number of potential impacts: it can undermine trust in the group or organization, reducing their impact, and undermining funding; can lead to attacks against the groups served by CSOs, particularly marginalized communities, often leading to political disempowerment and – in the worst cases – loss of life; and, finally, issue or group focused civil society groups often get caught up in disinformation campaigns designed to discredit or undermine their agendas, even if they are not attacked directly. As such, every civil society organization – regardless of its focus – is impacted by disinformation and has a role to play in combating it.

In addition to those civil society groups and interventions explicitly working on disinformation, the democracy assistance community must work with civil society writ large to ensure that they are prepared for information attacks designed to discredit an organization, its beneficiaries, or the issue area they work on.



HIGHLIGHT

RESEARCH FOCUS: RETALIATION AGAINST COUNTER- DISINFORMATION INITIATIVES

Beyond threats associated with disinformation campaigns targeting civic groups, perpetrators of disinformation also target organizations working to fact-check statements, identify narratives, and/or build public awareness of the issue of disinformation. Respondents to CEPPS interview research in Ukraine noted several instances of retaliation against civic groups working on disinformation, ranging from

public rebuttals and rhetorical attacks to harassment, physical threats, and vandalism.

That preparation should include:

- All civil society groups should be trained in basic data protection and information security to ensure that sensitive financial information, interior workings, and – most critically – membership databases or communications with vulnerable groups and individuals remain secure.
- Civil society groups should be encouraged to have a crisis response plan for information attacks. Who needs to be involved in response discussions? In what instances would the civil society group respond? How quickly will they respond? How will they ensure that a response reached the target audiences? Will beneficiaries or member groups be notified of information attacks or data breaches? How?
- Groups working on issues likely to be subject to disinformation should be trained in how to anticipate, identify, report, and counteract disinformation. Rapid response grants and capacity building initiatives should be put in place around specific issue areas.

BUILDING CIVIL SOCIETY CAPACITY TO MITIGATE AND COUNTER DISINFORMATION

9. INTERNATIONAL COLLABORATION (/TOPICS/CSOS/9-INTERNATIONAL- COLLABORATION)

International cooperation is a critical factor behind civil society success. In addition to the leverage issue vis-à-vis companies discussed earlier in the chapter, international cooperation allows civil society to share best practices in the rapidly evolving fields of digital forensics and counter-messaging, and to share information about emerging transnational threats and the proliferation of disinformation toolkits used by malign actors both foreign and domestic.

For instance, as COVID-19 took root, a coordinated Chinese Communist Party (CCP) information operation proliferated that was designed to sow misinformation about the origins of the virus, to undermine the successes of democratic actors in combatting the virus, and to amplify stories around CCP aid to countries struggling to contain and treat the virus. IRI convened a group of over a hundred representatives from civil society from every corner of the world to facilitate

information-sharing about CCP tactics and narratives related to the virus, as well as best practices for countering that information operation. Such networks and information-sharing are absolutely critical to civil society as they attempt to stay ahead of information threats.

Regional collaboration has also helped to expose and counter coordinated cross-border information operations. Activists in countries impacted by Russian disinformation have collaborated to share information about Russian tactics and narratives that are repeated across their countries, or where the same assets (accounts, pages, groups, content farms, etc.) are used across borders. They have also collaborated in applying open source intelligence (OSINT) to expose Russian lies: the [InformNapalm \(https://informnapalm.org/en/\)](https://informnapalm.org/en/) group is a volunteer effort comprised of individuals from across ten countries who expose “[evidence of Russian aggression to the world \(http://informnapalm.rocks/\)](http://informnapalm.rocks/)”, including [publishing the names of Russian servicemen \(https://www.the-american-interest.com/2020/06/02/activists-against-digital-lies/\)](https://www.the-american-interest.com/2020/06/02/activists-against-digital-lies/) who have fought in Ukraine, Georgia, and Syria based on the social media activity of those individuals.

As mentioned, the Poynter Institute's International Fact-Checking Network provides a mechanism for the certification of fact-checking groups according to its [principles \(https://ifcncodeofprinciples.poynter.org/\)](https://ifcncodeofprinciples.poynter.org/), and for coordinating fact-checking globally. [In addition, IFCN's system and members have been integrated into Facebook's online systems for reviewing and potentially downgrading content within it. \(/topics/platforms/3-efforts-promote-resiliency-digital-literacy-and-stronger-community-responses#IFCN\)](#) This has the potential for amplification both through the online tech platform and through the network of organizations sharing best practices and performing research and fact checks globally.

Some of the most successful civil society initiatives combatting disinformation are volunteer-run initiatives. This reflects a grassroots reaction to what is a relatively novel threat. However, online disinformation is not only here to stay, it is likely to metastasize and evolve as platforms, actors, and tactics proliferate. Civil society thus needs a funding model that recognizes the requirement for long term, dedicated, expert staffing. [Per Thomas Kent \(https://www.the-american-interest.com/2020/06/02/activists-against-digital-lies/\)](https://www.the-american-interest.com/2020/06/02/activists-against-digital-lies/), “Grants often fall in the \$10,000-\$50,000 range—hardly enough to hire staff and get major projects underway. Real breakthrough projects might be big-ticket items like opening radio and television stations to compete with broadcasters controlled by authoritarian governments and corrupt financial interests. Projects of this scope are almost impossible given the way funding is handled now.”

BUILDING CIVIL SOCIETY CAPACITY TO MITIGATE AND COUNTER DISINFORMATION

10. CONCLUSION AND RECOMMENDATIONS (/TOPICS/CSOS/10-CONCLUSION-AND- RECOMMENDATIONS)

Civil society plays a critical and multifarious role in information integrity infrastructure, but most organizations operating in this space are under-resourced, low capacity, and otherwise nascent. Funders and implementers need to invest in the long-term development of expertise at the grassroots level, in international collaboration, and in local to global communication in order to ensure that future threats to information integrity are dealt with promptly, and to create a global environment in which disinformation becomes a less effective tactic for hybrid warfare, political competition, or malign interventions in civic discourse.

Recommendations

Civic organizations play a key role in identifying and responding to information disorders, especially where they can establish reputations as relatively independent, objective actors. However, these advantages come with tradeoffs, especially if their constituencies tend to be relatively urban, highly educated, wealthier, or more internet-connected on average. Program designs should take care to target interventions to encourage uptake among underserved groups.

Network and coalition approaches to countering disinformation, including international collaboration, can identify comparative advantages, increase scale, and improve the diversity of programmatic approaches.

Relatedly, programs focused on civil society should incorporate an intentional focus on inclusion, and more specifically, intersectionality, particularly in coalition and network approaches. Support for civic groups should incorporate a distinct analysis to identify unique challenges faced by groups with intersectional identities within a specific historical context, since perpetrators of disinformation campaigns may rely on the apathy or complicity of non-marginalized identity groups. Collective action is more likely when these groups and individuals that are not politically marginalized understand that they have an interest in defending the rights of smaller and more vulnerable groups.

Civic organizations may consider partnering with existing political or social institutions to scale programmatic responses to disinformation, especially if the organization itself has a small or narrow audience. One example might include partnering with school systems to implement media-literacy programs.

Programs working on advocacy, especially around internet or platform regulation should consider the specific cultural context of debates surrounding tradeoffs between free expression and security.

Programs working with civic organizations to implement counter-disinformation programs should consider dedicated security training components, including cybersecurity, data protection, response plans for information attacks, and physical security from retaliation.

ELECTION MANAGEMENT BODY APPROACHES TO COUNTERING DISINFORMATION

0. OVERVIEW - EMB APPROACHES (/TOPICS/EMBS/0-OVERVIEW-EMB- APPROACHES)

Written by Lisa Reppell, Global Social Media and Disinformation Specialist at the International Foundation for Electoral Systems Center for Applied Research and Learning

Digital disinformation is a real and immediate threat for election management bodies (EMBs) around the world. However, election authorities in different countries are embracing to varying degrees the expectation that they have a substantive role to play in countering disinformation related to electoral processes. Some EMBs have sophisticated social media monitoring capabilities and dedicated teams to track and respond to disinformation; others do not have any social media presence at all. For all of them, disinformation is an unwieldy threat that is being brought to their door, while the immense, primary task of the EMB – administering credible elections – continues to be just as complex an endeavor as ever.

An EMB's resistance to taking up a role in countering disinformation may be based on an assumption that any response would require the institution to invest in a wholly new technical approach that pushes them beyond their legal, budgetary or human capacity. Though technology and social media have heightened the urgency and awareness of disinformation as a challenge to democratic processes and institutions, it is important to recognize that responses do not necessarily have to be technological in nature. **In addition to technology-forward responses that some EMBs may be equipped to adopt, there are also a range of responses that EMBs can take that build on existing core functions of public relations, communication and voter**

education. Finding an alternate way to frame an EMB's counter-disinformation efforts, such as investment in election authorities' crisis and strategic communication capacity, may also be a way to gain institutional support for new initiatives.

An EMB's specific role in contributing to the integrity of the information space around elections will vary based on its institutional mandate, resources, and capacity. Nonetheless, EMBs around the world are independently developing responses to counter disinformation in the electoral process and sharing lessons learned with peers. This section of the guidebook presents a global overview and preliminary analysis of the various EMB responses taken to counter electoral disinformation. The purpose of this analysis is to support election authorities as well as donors and implementers to combine, scale and adapt approaches based on an EMB's capacity and the unique context in which it is working.

"We manage not just the election – but there is another thing we have to be concerned about. This is the social media issue. This makes a very big noise, but it's not directly an election issue." – Commissioner Fritz Edward Siregar, The General Election Supervisory Agency of Indonesia (Bawaslu)

INFORMATIVE VS. RESTRICTIVE APPROACHES TO COUNTERING DISINFORMATION

A fundamental tension at the heart of how EMBs choose to respond to electoral disinformation is whether to focus on increasing dissemination of credible information or on restricting or sanctioning content or behaviors deemed problematic. While it may be possible to do both with adequate resources, for some EMBs it is a question of what the guiding principle behind their approach will be. In a report summarizing their disinformation efforts in 2018 and 2019, the National Electoral Institute of Mexico (INE) sums up this choice, and the philosophy behind their approach:

"Disinformation strategies challenged INE with the need to find a way to counter them. One alternative could have been undertaking a regulatory stance ... and punish[ing] pernicious practices; although it might have resulted in undue restrictions on freedom of expression. The other was to counter disinformation with its contrary: detailed, timely, and truthful explanation of the electoral process, its stages, tempos, stakeholders, and those in charge.... It was always clear for INE that this second option was the most adequate...."¹

Other EMBs, often in concert with a broader intra-governmental approach, error on the side of restricting content and behaviors as a means to prevent harms.

PROACTIVE, REACTIVE AND COLLABORATIVE STRATEGIES

The EMB strategies to combat disinformation discussed in this section of the guidebook are grouped into three categories: *proactive*, *reactive* and *collaborative*. Users can click on each strategy in the table below to explore global examples as well as analysis regarding what considerations should be made when choosing an approach.

EMBs can adopt **proactive strategies** in advance of electoral periods to promote trust and understanding of electoral processes, put contingency plans in place for when challenges emerge, and establish norms and standards for conduct during elections. Proactive strategies are more likely to build on pre-existing functions within an EMB. In designing a counter-disinformation strategy, EMBs and partners should acknowledge that *reactive* approaches that attempt to mitigate the impacts of disinformation once it is already in circulation can only address part of the problem. **Election authorities, donors and implementers should not let a bias toward technologically innovative programming undercut continued investment in building the types of durable capacity that make EMBs more resilient when disinformation challenges arise.**

EXPLORE PROACTIVE STRATEGIES:

1. Strategic Communication and Voter Education to Mitigate Disinformation Threats (/topics/embs/1-strategic-communication-and-voter-education-mitigate-disinformation-threats): *Building resilience to misinformation and disinformation by ensuring voters receive credible information early, often, and in ways that resonate with them.*
2. Crisis Communication Planning for Disinformation Threats (/topics/embs/2-crisis-communication-planning-disinformation-threats): *Putting systems and processes in place so that an EMB is prepared to rapidly and authoritatively respond to misinformation and disinformation in high-pressure situations.*
3. EMB Codes of Conduct or Declarations of Principle for the Electoral Period (/topics/embs/3-emb-codes-conduct-or-declarations-principle-electoral-period): *Creating norms and standards for political parties, candidates, media and the electorate at large that promote the integrity of the information environment around elections.*

Reactive strategies to track and respond to messages in circulation that have the potential to disrupt electoral processes, generate distrust in elections, or illegitimately shift electoral outcomes are an important aspect of countering disinformation. **Reactive interventions may be the first to come to mind in designing a counter-disinformation approach, but these approaches can be the most technologically difficult for EMBs to implement and the most resource**

intensive. While reactive interventions are an integral part of a multifaceted response to disinformation, combining them with proactive strategies and ensuring that an EMB has the capacity and appetite to implement them effectively are critical for ensuring an effective approach.

EXPLORE REACTIVE STRATEGIES:

4. Social Media Monitoring for Legal and Regulatory Compliance (/topics/embs/4-social-media-monitoring-legal-and-regulatory-compliance): *Monitoring social media during electoral periods to provide oversight of the social media use of candidates, campaigns and the media.*
5. Social Listening to Understand Disinformation Threats (<https://counteringdisinformation.org/topics/embs/5-social-listening-understand-and-respond-disinformation-threats>): *Distilling meaning from conversations happening online in order to inform EMB messaging and responses to misinformation and disinformation during electoral periods.*
6. Disinformation Complaints Referral and Adjudication Process (/topics/embs/6-disinformation-complaints-referral-and-adjudication-process): *Establishing a mechanism or process by which election authorities or election arbiters can adjudicate and remedy instances of disinformation.*

Regardless of how narrowly or broadly an EMB interprets its mandate to engage in counter-disinformation work, to achieve maximum impact the efforts of election authorities must be **coordinated** with the efforts of other state agencies and institutions. EMBs are likely to maximize the impact of their efforts through coordination or exchange with other stakeholders, including social media and technology companies, civil society and traditional media actors, as well as other state entities. There will always be aspects of the disinformation problem that fall outside the mandate of the EMB. **The appropriate allocation of responsibilities in a way that allows EMBs to focus their counter-disinformation efforts on electoral integrity considerations – while coordinating their response with other stakeholders better equipped to handle other facets of the problem – will enable a more concentrated and focused effort on the part of the EMB.**

EXPLORE COORDINATION STRATEGIES:

7. EMB Coordination with Social Media and Technology Companies (/topics/embs/7-emb-coordination-technology-and-social-media-companies): *Coordination between EMBs and technology and social media companies to enhance the dissemination of credible information or restrict the spread of problematic content during electoral periods.*
8. EMB Coordination with Civil Society and Media (/topics/embs/8-emb-coordination-civil-society): *Partnerships with civil society and media to build coalitions to counter disinformation and enhance an EMB's ability to monitor and respond to misinformation and disinformation.*
9. EMB Coordination with Other State Agencies (/topics/embs/9-emb-coordination-other-state-entities): *Partnerships with other state entities to distribute responsibilities and coordinate*

responses to misinformation and disinformation.

10. Peer Exchange Among EMBs on Counter-Disinformation Strategies (/topics/embs/10-peer-exchange-among-embs-counter-disinformation-strategies): *Creating opportunities for exchange of lessons learned and good practice among election authorities.*

SHOULD EMBS HAVE A RESPONSIBILITY TO COUNTER DISINFORMATION?

This is a question on which EMBs are not in agreement. Differences in legal mandates, political context, availability of resources, and technical capacity all influence the degree to which an EMB might be willing and able to adopt a substantive role in countering disinformation.

Different EMBs highlight various aspects of their **legal mandate** to justify their role in counter-disinformation work. Oversight of the conduct of political candidates or the media during the electoral period, or a voter education or voter information mandate, are some of the avenues that EMBs might use to define the parameters of their role in countering disinformation. A broad responsibility for EMBs to maintain the fundamental right of citizens to vote can also be grounds for an EMB to take an active role. Differing legal mandates will inform what programming is possible to implement with an EMB. For instance, an extension of some EMBs' mandates to monitor traditional media during electoral periods might naturally be extended to monitoring social media as well. For other EMBs, the monitoring of social media during electoral periods would be an inappropriate overstep of their legal mandate. Any programming to bolster EMB-responses to disinformation must be grounded in a thorough understanding of the bounds of what is legally permissible.

From a **resource perspective**, strained budgets or limited control by the EMB over how to use allocated funds can make it difficult to dedicate resources to counter-disinformation activities, particularly if they are seen to divert resources from other essential aspects of election administration. If EMBs struggle to muster the resources to conduct their core mandate of delivering elections, the investment of resources to build out a significant capacity to address disinformation is likely to be untenable.

From a **technical and human capacity perspective**, EMBs may also lack the human resources to contemplate responses to disinformation that are time intensive or technologically sophisticated. Recruiting and retaining staff that have knowledge of social media and technology more broadly can be difficult, particularly if the EMB is attempting to build out an entirely new capacity, as opposed to strengthening or investing additional resources in a capacity that already exists.



HIGHLIGHT

While disinformation responses can be housed within different departments of an EMB, many EMBs have chosen to give this mandate to the public relations or

As a final consideration, the political context in which an EMB operates may also impact the **institutional independence** of the EMB in ways that limit its efficacy as a counter-disinformation actor. In instances where an EMB's actions are subject to or constrained by the political influence of domestic actors, extending an EMB's mandate to counter disinformation may be ineffective and the EMB may be reluctant to take on such a role. If the EMB is already perceived to be partial, its efforts to counter disinformation may also further damage its credibility in the eyes of the public.

communications staff. The Independent National Electoral Commission of Nigeria, with 90 full time communications staff, has enacted and can consider counter-disinformation approaches that are unlikely to be practicable for an EMB with a communications staff of only a few people.

ELECTION MANAGEMENT BODY APPROACHES TO COUNTERING DISINFORMATION

1. STRATEGIC COMMUNICATION AND VOTER EDUCATION TO MITIGATE DISINFORMATION THREATS (/TOPICS/EMBS/1-STRATEGIC- COMMUNICATION-AND-VOTER- EDUCATION-MITIGATE- DISINFORMATION-THREATS)

In an era of information overload and digital disinformation, it is critical that EMBs are able to cut through the noise with proactive and focused messaging. As credible information can easily be lost in a sea of distracting, problematic and misleading messages, the impetus is on authoritative actors – such as EMBs – to ensure credible messages are reaching the right audiences in ways that resonate with them. Proactive counter-disinformation messaging can be embedded within an EMB's larger communication strategy, or can be one of the outputs of a dedicated counter-disinformation planning process. Either way, an effective communication strategy requires planning and refinement in advance of an election. Depending on patterns of social media use in their country, an EMB may also have the opportunity to use social media to reach specific audiences susceptible to or likely to be targeted by disinformation, such as women, people with disabilities, and people with lower levels of education, among other groups.

Like all of the measures in this section of the guidebook, proactive communication strategies can and should be combined with other proactive, reactive or coordinated responses to form a comprehensive and inclusive approach to improving the integrity of the information environment around elections. The balance or combination of these measures is likely to vary from one election to the next.

Proactive messaging should not be confused with the more limited idea of messaging that raises public awareness about the existence of disinformation. Awareness of disinformation as a threat is already on the rise, with a [2018 Pew Center survey](https://www.pewresearch.org/internet/2019/05/13/publics-in-emerging-economies-worry-social-media-sow-division-even-as-they-offer-new-chances-for-political-engagement/)

(<https://www.pewresearch.org/internet/2019/05/13/publics-in-emerging-economies-worry-social-media-sow-division-even-as-they-offer-new-chances-for-political-engagement/>) showing that almost two thirds of adults across 11 surveyed countries believe “people should be very concerned about exposure to false or incorrect information.” This finding is further supported by CEPPS public opinion research.¹ Messaging can and should seek to raise awareness of the need to be critical of information sources, think before sharing content, and other basic tenets of digital literacy. However, messaging should also focus on the broader goal of communicating in ways that build trust (https://behavioralpolicy.org/wp-content/uploads/2017/05/BSP_vol1is1_Schwarz.pdf) in the EMB and faith in the integrity of electoral processes.

1.1 EMB VISIBILITY HAS VALUE

Building a track record of consistent communication can help an EMB to message with authority during times of confusion or heightened tension that might stem from mis- or disinformation. As a new wave of digital disinformation has made clear, investments in an EMBs’ capacity to deliver their core communication mandate through new and established channels is increasingly vital.

The INE of Mexico provides one model to consider for EMBs’ developing their own organizational approaches to countering disinformation. INE designed a robust digital media strategy ahead of 2018 elections, aiming to increase the volume of credible, engaging content contending for user-attention on social media. During the 2018 electoral period, INE produced and disseminated over six thousand pieces of digital content, which were also available through a centralized website (<https://centralectoral.ine.mx/>), focused on public outreach.²

“We bet on a different strategy – to confront disinformation with information.” — Dr. Lorenzo Córdova Vianello, Councilor President of the National Electoral Institute of Mexico

The INEC of Nigeria deploys its longstanding institutional investment (<http://www.elections.org.za/SEIDA2020/Documents/Dr%20Sa'ad%20Umar%20Idris%20INEC%20So>) in public communication as a bulwark against disinformation. During electoral periods, the INEC provides daily televised briefings, participates in live TV interviews, issues regular press

statements to explain the policies and decisions of the commission, and runs the INEC Citizens Contact Centre (ICCC) to provide the public with access to the commission and communicate with critical stakeholders. INEC has also had an active social media presence for more than a decade, using it as a channel to disseminate information and interact with voters. As the INEC confronts digital disinformation, their existing communication capacities are being reconsidered and adapted to enhance INEC's transparency, credibility and perceived integrity in order to sustain public trust and confidence.

The Brazilian Superior Electoral Court (TSE) augmented their traditional public outreach strategies through investment into widely-adopted mobile applications that allow election authorities to communicate rapidly and directly with voters and poll workers. The “e-Título (<https://apps.apple.com/us/app/e-t%C3%ADtulo/id1320338088>)” mobile app works as a virtual voter ID card, helps voters identify their polling stations and provides an avenue for direct communication between the TSE and voters. The “Mesários (https://play.google.com/store/apps/details?id=br.jus.tse.eleitoral.mesarios&hl=en_US&gl=US)” application provides information and training to poll workers. During the 2020 electoral period, more than 300 million messages were sent to the almost 17 million users of these apps with timely and reliable information on election organization, health protocols amidst Covid-19, and tips to fight fake news.



Nigeria's EMB has an active social media presence, using it as a channel to disseminate information and interact with voters.

PHOTO: FACEBOOK PAGE OF Independent National Electoral Commission (Nigeria)

“We want to prevent the dissemination of fake news not with content control, but with clarification, critical consciousness and quality information.” — Justice Luís Roberto Barroso (<https://www.tse.jus.br/imprensa/noticias-tse/2020/Julho/ministro-luis-roberto-barroso-se-reune-com-parceiros-no-combate-a-desinformacao>), President of the Brazilian Superior Electoral Court (TSE)

1.2 COUNTER THE OBJECTIVES OF THE PROPAGANDA, RATHER THAN THE PROPAGANDA ITSELF

A proactive communication strategy will attempt to anticipate what categories of false or problematic messages are likely to gain traction and be damaging during a specific election, and will then aim to build resilience in those areas. The Harvard Belfer Center's [Handbook on National Counter Information Operations Strategy](#)

(<https://www.belfercenter.org/sites/default/files/files/publication/CounterIO.pdf>) emphasizes that communicators should seek to counter the *objectives* of propaganda, rather than the propaganda itself. A proactive communications campaign that builds public understanding of election procedures and public trust in the integrity of the EMB is likely more effective preparation than trying to anticipate each false narrative malign actors might choose to employ, particularly since these actors can change and adapt strategies quickly. If one false narrative is not gaining traction, they can simply switch to another.

“Given the volume and content of information operations that competitors can spew out through social and traditional media, [authorities] cannot and should not respond to each false narrative individually. Addressing the content directly adds fuel to the narrative’s fire.”
— *Belfer Center Handbook on National Counter Information Operation Strategy*

Electoral disinformation within an EMB’s purview might seek to undermine faith in the value or integrity of elections or election authorities, incite electoral violence, or seed suspicions of fraud that lay the groundwork for post-electoral legal challenges. As a proactive approach, EMBs and other stakeholders could design a communication strategy in advance of the election around the goals of enhancing transparency and building understanding of electoral processes, highlighting election security measures, or explaining the election dispute resolution process.

Electoral disinformation might also seek to prevent specific groups from participating in the electoral process by spreading false information about the rights of certain groups and by targeting specific groups with false election information. Disinformation campaigns frequently manipulate and amplify hate speech and identity-based social divisions, allowing malign actors to heighten social polarization for personal or political gain. EMBs can proactively combat these efforts by ensuring that their communications strategies target majority groups and minority groups with messages that highlight the rights of women, people with disabilities, and other marginalized groups to equally participate in the electoral process as well as other targeted voter information. To reach different groups, information might need to have unique dissemination strategies that differ from general voter education efforts – person-to-person; in markets, churches, and other common places; in simple language, in images, or in minority languages – to take into account barriers these groups face when accessing information.

Given changes to the administration of elections introduced through electoral reform in 2014 in Mexico, misunderstanding of the new processes was a potential source for misinformation and disinformation during the 2018 election process, the first under the new reforms. A key push of INE’s public communications strategy ahead of the elections was to build understanding of the

mechanics of voting, counting and results transmission by explaining new processes clearly and simply so that people knew what to expect at every moment during the election. Communicating in a way that reinforced INE's political neutrality was also key, as the authorities knew that partisan or bad actors might attempt to politicize the institution.

In Indonesia, where intercommunal fault lines are ripe for exploitation, the election oversight body, Bawaslu, created PSAs against incitement to violence and hate speech (<https://www.instagram.com/p/Bu2zVOqFbJK/>) and promoting digital literacy (https://www.instagram.com/p/BwJPwWll_l4/) in advance of 2018 elections. The PSAs were developed with IFES support and disseminated via YouTube, Instagram, Facebook, Twitter, WhatsApp, and Bawaslu's websites as well as digital billboards throughout Jakarta. These PSAs were followed by a second round focusing on the role of participative public election monitoring and tutorials on election violation reporting tools as well as cautions against incitement to violence and disinformation.³ The strategy and storylines for the PSAs were developed through a consultative workshop facilitated by IFES with both election management bodies and key civil society partners.



HIGHLIGHT

Indonesia has two distinct election management bodies. The Komisi Pemilihan Umum (KPU) which administers elections in Indonesia as well as the election supervisory body, Badan Pengawas Pemilihan Umum (Bawaslu) which is charged with monitoring and oversight of the electoral process.

CEPPS conducted fieldwork in Jakarta in late 2019 to inform the development of this guidebook.

FEATURED INTERVENTION COUNTERING INFORMATION INFLUENCE ACTIVITIES: A HANDBOOK FOR COMMUNICATORS (/INTERVENTIONS/COUNTERING-

INFORMATION-INFLUENCE- ACTIVITIES-HANDBOOK- COMMUNICATORS)

Sweden's Civil Contingencies Agency (MSB) developed a handbook and training delivered to municipal election administrators in Sweden's decentralized election system.

1.3 EFFECTIVE MESSAGING TO PROMOTE INFORMATION INTEGRITY

Make Messages Engaging

In the face of constant innovation in communication methods, EMBs must respond to the evolving nature of communication. By no means does this mean that EMBs should abandon traditional communication channels; radio, television and newspapers still directly reach a larger share of the population than social media in most countries, and traditional media is still a part of the information ecosystem that amplifies false and misleading information that originates online. However, revising and innovating within their communication approaches can help EMBs meet their key audiences with messages they will more readily consume and remember. Explicitly identifying ways to create engaging content can be an important part of an EMB counter-disinformation strategy.

Even if an EMB is already using social media to some degree, strategic consideration should be given to the value of engaging with voters on new platforms or utilizing new features on the platforms where they have an established presence. For example, though the South African IEC



HIGHLIGHT



To counter hate speech and the spread of disinformation, in partnership with CEPPS/IFES, the Union Election Commission (UEC) of Myanmar developed animated public service announcements (<https://m.facebook.com/watch/?v=834197684009924&rdr>) that were shared on the UEC and partner social media channels and websites. This was also adapted to a comic book (<https://merin.org.mm/en/publication/finding-and-sharing-accurate-information>) and translated into 20 ethnic languages.

has been present on Facebook and Twitter for nearly a decade, during 2019 they made use of a voter registration Snapchat feature for the first time. This in-app feature connected Snapchat users to voter registration resources, and the number of South African users taking advantage of this feature to register exceeded averages from other countries.⁴ Brazil's TSE, while continuing to expand their use of Instagram, Facebook and Twitter, established a TikTok presence less than two months before 2020 local elections. Given that content on TikTok can organically reach large audiences without needing to build a follower base first, in those two months, the TSE's TikTok account gained 20,000 followers and millions of views for their library of approximately 80 videos; a TikTok video outlining health protocols to be followed on Election Day achieved over 1.2 million views alone.

In Taiwan, the form of counter messages coming from official channels is encouraged to be funny and "memetic" to increase the likelihood that counter messaging can organically go viral via the same channels through which disinformation proliferates. For example, to prevent the transmission of misinformation and disinformation during the COVID-19 pandemic, Digital Minister Audrey Tang has established the Taiwan FactCheck Center (<https://www.atlanticcouncil.org/blogs/new-atlanticist/lessons-from-taiwans-experience-with-covid-19/>), which include Meme Engineering teams that partner with national comedians to clarify online rumors to the public in an expedient, humorous, and effective way. This 'humor over rumor' (<https://www.businessinsider.com/taiwan-coronavirus-strategy-digital-campaign-dog-mascot-2020-6>) strategy is acknowledged as a



HIGHLIGHT



The KPU in Indonesia created 3,000 anti-hoax memes

(<https://www.kpu.go.id/index.php/pages/detail/>) which consisted of infographics and other branded social media content in advance of the election. Content created by the central KPU would be modified by regional offices in response to local context and translated into local languages.



HIGHLIGHT

Premier Su Tseng-chang shared this image on Facebook showing him as a young man with a full head of hair, as means to dispel online misinformation of new government

critical strategy in helping curb the spread of COVID-19 in Taiwan, and this approach can be adapted when countering disinformation beyond the pandemic.

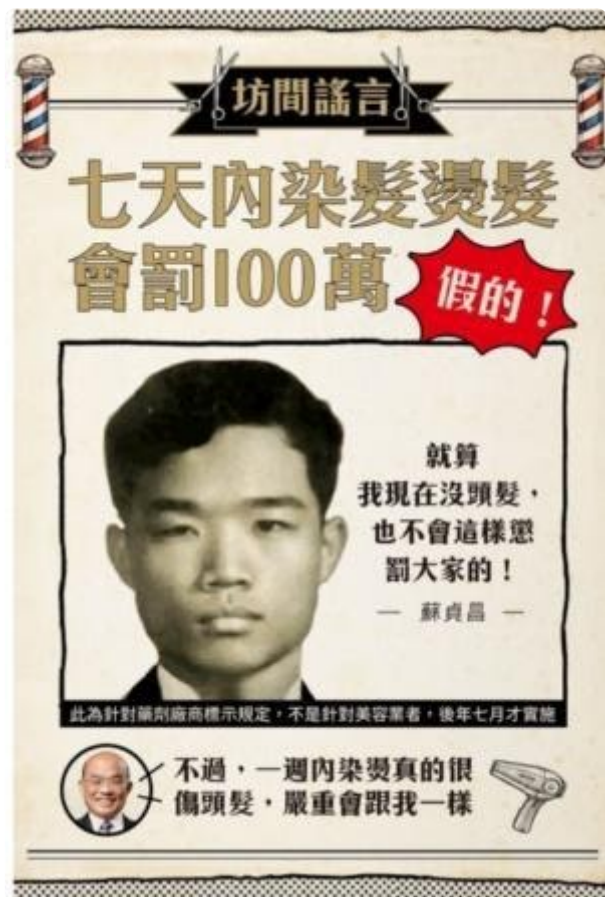
To make content both engaging and credible, EMBs can also identify trusted messengers with the ability to reach specific audiences. Establishing lines of communication with leaders or members of religious groups, sports clubs, libraries, professional networks or other traditionally apolitical spaces might be a means to reach new audiences with proactive messaging. Nigeria's INEC, for example, works with actors and other celebrities to visit college campuses and build enthusiasm among youth voters. Brazil's TSE partnered with football clubs as part of their #NaoTransmitaFakeNews campaign urging users to not spread fake news. Eighteen football clubs participated in the campaign, which garnered more than 80 million Twitter impressions across the first and second round of the election.

These networks of trusted messengers, when built in advance, can also be used as dissemination channels and amplifiers in instances where false information needs to be debunked, an approach that is discussed further in the subsection on crisis

communication. While these networks can be built by national election authorities, regional EMBs might also benefit from building their own subnational networks of trusted messengers.

Make Messages Inclusive and Accessible

Ensuring that messages are inclusive and accessible to all people and, in particular, groups that have been historically marginalized, is a key consideration for EMBs. EMBs should ensure that they consider the diverse ways people access voter information. For example, men in a given country might be more likely to rely on television for voter information, while women might rely on radio messages or conversations with neighbors. EMBs can conduct surveys or polls, or consult with organizations that represent people from different marginalized groups, in order to understand how different voters access information and then be responsive to those needs.



regulations on hair salons. It includes the mock caution: 'Dyeing and perming within seven days really damages your hair, and in severe cases you'd end up like me.'

PHOTO: FACEBOOK PAGE OF PREMIER SU TSENGCHANG

In addition, many social media platforms offer ways for users to easily add accessibility features, such as alternative (alt) text⁵ to describe photos for screen readers, posting a transcript for an audio file such as a radio recording, or including subtitles or captions⁶ for videos. EMBs that use these features, and that include actors and images of people with disabilities and other diverse identities in their campaigns make their content more inclusive and accessible. EMBs can help ensure that the content they produce is accessible by distributing messages in multiple formats, such as sign language, easy-to-read, and local languages, and consulting with civil society organizations on the most commonly used platforms, pages and handles.

For example, ahead of the August 2020 elections in Sri Lanka, the EMB collaborated with a group of DPOs to create a social media campaign to ensure people with psychosocial disabilities knew they had the right to vote and to raise awareness with political parties of the need to eliminate derogatory language from their political campaigns. The campaign, produced both Sinhala and Tamil, reached nearly 50,000 people and resulted in the EMB Chairman releasing a public statement acknowledging the political rights of people with psychosocial disabilities.

Another key point in accessibility is considering the gap in access to and knowledge of certain technologies for certain populations. The gender gap in access to technical tools and the internet, for example, is well-documented and underscores the need to continue to disseminate messages in ways that are accessible to those who might not have consistent access to technology.

Make Messages Memorable

To make a proactive counter narrative memorable in the face of an onslaught of repetitive and reinforcing disinformation, it must have a clear point and it must be repeated many times. Research suggests that for both true and false claims, information that is repeated (<https://theconversation.com/unbelievable-news-read-it-again-and-you-might-think-its-true-69602>) feels more true, even if it goes against what you think you know.

In response to a fraudulent campaign in which bad actors were using the Central Election Commission's (CEC) name to knock on doors and collect personal information, the Georgian EMB widely disseminated the message that it does not collect information in this manner, and then repeated the message via multiple communication channels. The EMB's response did not simply contradict the message that was being spread, but it used the initial incident to raise public awareness about the methods that were being used to deceive and to share a clear message on how to get credible information if voters were faced with similar uncertainty in the future.

Messaging does not need to be technologically groundbreaking to be effective, but adapting approaches to fit new needs is critical. Stretched resources and staff, outdated or nonexistent strategic communication strategies, and a belief that the truth of a message should speak for itself can undermine the communication effectiveness of EMBs. Reflecting on a press conference that her institution had held to debunk false information circulating about upcoming elections, a staff member of an East African Election Commission observed that it had only served to increase the virality of the rumors and encourage the disinformation. Reactive, static and unengaging counter-messages are less likely to achieve the desired result of building trust in the process.

1.4 TAKE ADVANTAGE OF UNIQUE ASPECTS OF SOCIAL MEDIA FOR EMB USE

Social media can provide EMBs that are equipped to use it with a potent tool for increasing institutional transparency, building trust, and executing their voter education mandates. While institutional use of social media is no longer a cutting-edge idea, there are EMBs that still do not use social media at all, and many that are working to keep their approaches current as patterns of social media use evolve. For countries with high rates of social media penetration, investment in an EMB's social media capacity is a moderate cost, high impact way of reaching key audiences and providing a counter narrative on the same channels where digital disinformation is originating and spreading.

Use social media for two-way communication

Social media has the potential to provide a direct channel of dialogue between EMBs and voters. Training and empowering designated EMB staff to take advantage of this two-way channel for communication is therefore very important. Because of the informality of the medium, social media has the potential to be a more authentic, open, timely and responsive means of communication. An EMB's willingness to directly engage with voters through their social media channels to provide quick, personal communication can build trust and provide an authoritative source where voters can seek or verify information. In deciding to adopt a more robust social media presence, EMBs should be resourced and prepared to follow through on this potential. Once an EMB opens this channel for conversation, they must be ready to sustain it.

"The deployment of Social Media as a communication strategy employed by INEC has had a profound impact on electoral processes, changing the channels used by citizens and voters to obtain information from the traditional media or one-way communication channel to the mobile-based platforms that allow for two-way interactions through user-generated content and communication." — Dr. Sa'ad Umar Idris, Director-General INEC Electoral Institute, Nigeria

Segment audiences and reach target audiences

Social media also allows for the potential to segment and reach audiences with messages more uniquely calibrated to resonate with them. This is a powerful strategy already employed by disinformation actors.

There are two lenses to use when identifying audiences that an EMB may want to target with specialized counter-disinformation messaging. Considering both of these at the outset of developing a counter-disinformation communication strategy can yield different insights into which audiences to reach and how to reach them.

The first lens is to consider audiences that are likely to be *consumers* of misinformation and disinformation that might impact their willingness or ability to participate. For example, an EMB might identify first time voters, voters with disabilities, voters from an ethnic or linguistic minority - or any other group of voters -- as particularly at risk of encountering disinformation designed to suppress their democratic participation. By identifying tactics that might be used to inhibit the participation of these groups, EMBs can design and target content that dispels misunderstanding about voter registration, builds understanding of the accessibility of polling stations, or outlines the steps taken to ensure the secrecy and safety of casting a ballot. It is important, of course, for the EMB to understand if these targeted populations are actually using social media platforms (and which ones) before employing this strategy. For example, certain marginalized groups might be more likely to use specific social media platforms because of different individual, institutional, and cultural barriers.

The IEC of South Africa identifies youth, special voters and those voting abroad in their communication planning as distinct audiences they are trying to reach with specific messages. Furthermore, they build discrete communication campaigns into their overall communication strategy, including messaging around registration, applications for special voting and voting abroad, voting procedures and awareness building about digital disinformation. Integrating this segmentation into a cohesive communication strategy that includes social media can be an important way of proactively using social media to provide information, create a feedback loop, and reach audiences that might need more information as a precursor to participation.

The second lens is to think about audiences that might be the *subject* of a disinformation campaign. This might take the form of a disinformation campaign that evokes existing currents of hate against marginalized populations to suppress participation, allege electoral fraud, or promote outrage among dominant identity groups. For example, a disinformation campaign may be designed to intimidate women candidates into dropping out of a race or to allege that immigrant populations are engaging in large-scale voter fraud.

The complexity of this task can be tailored to match the needs and capacity of an individual EMB, recognizing that for some EMBs only very basic approaches will be possible or advisable and for other EMBs, advanced techniques would be entirely appropriate.

It should be noted that in the hands of commercial entities and malign actors, tactics such as those above have been understandably treated with suspicion. While EMBs should always adhere to a high standard of data privacy and data protection, these widely available tools are largely value neutral – it is the uses to which they are put that determine their ethical implications. If EMBs and other



HIGHLIGHT

Equipping EMBs to use social media to greater efficacy to reach different audiences could include:

- Using social media analytics to determine what types of content are performing well and which audiences are and are not being reached.

democratic institutions do not take advantage of the ways in which social media tools can be leveraged to promote democratic goals and the integrity of their institutions, then they can never hope to compete in their ability to shape the information space around elections or around democracy more broadly and will continue to be outmatched by bad actors on the messaging front.

- A/B message testing, which enables the content creator to compare the performance of different pieces of content so they can quickly pivot toward high performing messaging strategies while jettisoning underperforming content.
- Using the targeted advertising features of social media to reach defined audiences.

“You have to use social media to engage proactively. If you only use it to react, control or limit social media then that is a losing wicket” — Vice-Chairperson Janet Love, Electoral Commission of South Africa

ELECTION MANAGEMENT BODY APPROACHES TO COUNTERING DISINFORMATION

2. CRISIS COMMUNICATION PLANNING FOR DISINFORMATION THREATS (/TOPICS/EMBS/2-CRISIS-COMMUNICATION-PLANNING-DISINFORMATION-THREATS)

EMBs face a potent mix of pressures, including: heightened public perception of disinformation as a threat to elections; pressure on them to be seen actively countering disinformation; differing levels of understanding of the nature of the problem among EMBS; and the time-sensitive nature of effective responses. Given this context, an EMB's reaction in the moment might be informed by a perception of immediate need rather than reflecting a larger strategy best suited to promoting electoral integrity. Crisis Communication planning can create a roadmap for EMBS to respond to electoral disinformation during sensitive stages of the electoral process. In instances where an

EMB has historically relied on ad hoc communication strategies during a crisis, programmatic investment can help EMBs formalize a crisis communication strategy to improve the speed and accuracy with which they are able to respond to mis- and disinformation.

One tactic of disinformation campaigns – whether led by foreign or domestic actors – can be the deliberate attempt to create a crisis mentality in order to sow distrust or confusion and undermine faith in democratic institutions and the electoral process. Not all misinformation or disinformation is indicative of a crisis, and determining the timing and form of an EMB’s response, and in which circumstances it will make a response, is part of the preparatory work that can help an EMB focus resources and decision making when needed.

“Know the way you will react if a problem presents itself, if fake news comes out. The EMB can’t just receive hits.” — Dr. Lorenzo Córdova Vianello, Councilor President of the National Electoral Institute of Mexico

2.1 DON'T CREATE YOUR OWN CRISIS

Due to the added attention placed on digital disinformation, EMBs may feel compelled to respond to any and all items of election-specific misinformation or disinformation that they encounter. Crisis communication planning can help establish criteria for what circumstances will warrant a response from the EMB, and in what form. Highlighting the existence of a piece of false or misleading content for the purposes of rebutting it may not always be the best course. In explaining their thinking behind whether to debunk or ignore such content, nonprofit [First Draft writes](https://firstdraftnews.org/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-de%CC%81sinformation-1.pdf?x41819) (<https://firstdraftnews.org/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-de%CC%81sinformation-1.pdf?x41819>) that, “[i]f certain stories, rumours or visual content, however problematic, were not gaining traction, a decision was made not to provide additional oxygen to that information. The media needs to consider that publishing debunks can cause more harm than good, especially as agents behind disinformation campaigns see media amplification as a key technique for success.” This same consideration is important for EMBs. Crisis communication planning allows the time and space for EMBs to develop best practices for how they will provide clarifications or rebuttals so that they do not inadvertently exacerbate the problem. Good practice on how to provide effective fact checks continues to evolve, a topic that is explored further in the [guidebook subcategory on fact checking](/topics/csos/2-fact-checking) (</topics/csos/2-fact-checking>).

2.2 CREATE CLARITY ON LINES OF COMMUNICATION AND AUTHORITY

An integral part of crisis communication planning is ensuring that information flows and hierarchies are delineated in advance. This can be particularly relevant in instances when the EMB is called upon to confirm facts or issue clarifying statements and counter-narratives quickly. A clear and direct communication protocol to coordinate responses can be essential to restoring public confidence, as vague or conflicting clarifications can exacerbate the problem. It is imperative to have clarity on who has the right to issue statements and through what means those statements will be made – not only within the EMB, but in consultation with other state agencies that may be called upon to clarify. Failure to do this can add fuel to the very mis- and disinformation a public statement can be intended to quell. For example, in response to alleged out-of-country voting fraud taking place in Malaysia, the two different Indonesian electoral bodies initially issued contradictory clarifying statements (one claiming that the allegation was entirely fabricated, and the other stating that the issue was real, but minor and had already been detected) – which created more confusion and potentially undermined the credibility of both bodies. Having a clear and expeditious protocol in place for how the two agencies would coordinate messaging could have helped avert this misstep.

Crisis communication protocols should strike a balance between expedience and internal checks for accuracy. An EMB should avoid having communications choke points whereby requests for clarification cannot be responded to with speed. In the case of Indonesia, third-party fact-checking civil society organization MAFINDO would frequently call on the KPU and Bawaslu to issue a rebuttal of false or misleading information in circulation, but the CSO reported that response times varied significantly, at times taking several days to get a response, if one was received at all. Because the speed with which a false rumor is rebutted or removed once it has started to gain traction has a significant impact on the ultimate reach of that information, for cases where there is a clear-cut answer to be given, the right individuals should be given the power to clarify.

2.3 BALANCING MULTIPLE PRIORITIES

Crisis communication planning can also help to establish institutional guidelines on balancing communication priorities with other electoral priorities. “While it can be important for the public to see leaders pitching in during a crisis response, there is a limit.”¹ For example, Indonesian electoral authorities were very active in the investigation of cases of viral misinformation and disinformation in the run up to 2019 Elections. In the case of a rumor that cargo ships (<https://www.straitstimes.com/asia/se-asia/jakarta-probing-online-claims-about-containers-filled-with-ballots-for-president-jokowi>) full of pre-voted ballots had arrived in Jakarta, the commissioners themselves mobilized (<https://www.straitstimes.com/asia/se-asia/jakarta-probing-online-claims-about-containers-filled-with-ballots-for-president-jokowi>) to go to the port late in the evening on the day the rumor gained traction in order to investigate and issue a public statement. Similarly, a few days before the election, commissioners were deployed to Malaysia (<https://www.bloomberg.com/news/articles/2019-04-12/indonesia-probes-poll-fraud-in-malaysia-as-diaspora-casts-vote>) on short notice to rebut claims that fraudulent out-of-country voting (<https://www.bloomberg.com/news/articles/2019-04-12/indonesia-probes-poll-fraud-in-malaysia-as-diaspora-casts-vote>) was taking place. For a severe case in which a false claim has the potential

to derail the election, this response was transparent and visible, but a careful calculation should be made in terms of the best investment of time of EMB commissioners and staff, particularly in close proximity to elections when there are competing demands.

2.4 COORDINATION WITH OTHER STATE ENTITIES

The EMB may be the lead agency in the crisis response, or it may be one member of a network. Misinformation and disinformation are rarely siloed and clear-cut, and will often include aspects that are within an EMB's purview to rebut, such as false or misleading information directly related to the electoral process, in combination with other issues upon which another government agency might be better positioned to comment, such as public health concerns or rumors of violence on Election Day. The subcategory addressing EMB Coordination with Other State Entities (</topics/embs/9-emb-coordination-other-state-entities>) provides additional considerations on this topic.

Crisis communication planning must also include the post-electoral period. While an EMB may be active during the campaign period and on Election Day in monitoring and responding to problematic content, the period immediately following Election Day is one of the most at-risk periods for false and misleading information. Misinformation and disinformation that emerge during this period can have implications for public acceptance of the results or post-electoral violence if, for example, narratives of fraud or malpractice in polling, counting or results transmission gain popular traction in ways that leave citizens feeling disenfranchised. An interlocutor involved in media monitoring in South Africa notes that bad actors may modify their behavior for the better during campaign periods when they are aware of enhanced media monitoring and enforcement efforts, but observed that in South Africa "vile" content ramped up as soon as enhanced monitoring efforts ended.

The immediate post-electoral period can be a particularly strenuous period for EMBs as polls close and results are counted, aggregated and certified. Furthermore, the EMB is likely to be called on to resolve post-election complaints and they may also be a party in electoral cases that go to the courts. The coinciding of this exceptionally busy period for the EMB with a window of time that is particularly ripe for misinformation and disinformation means that a clear plan for communication protocol during the post-electoral period is essential, including clarity on communication with and shared responsibilities among state entities. Indonesian electoral authorities experienced this first hand (<https://www.thejakartapost.com/news/2019/05/22/post-election-unrest-grips-jakarta.html>) following 2019 elections when rumors of electoral fraud led to protests that resulted in the deaths of 9 people and the restriction of social media access by government authorities in an attempt to stem the spread of misinformation and unrest. In the run up to Election Day, Bawaslu had been playing a leading role in coordinating responses to electoral disinformation and flagging problematic content for removal by the social media platforms. However, strained capacity in the days following the election forced them to step back from this role with the expectation that other government agencies would be able to step in. Crisis

communication planning can help facilitate a smooth transition of responsibility from the EMB to other agencies in times where that is appropriate. A plan can help determine in advance in what way and at what point responsibilities might shift – both public communication responsibilities as well as communication with the social media platforms.

2.5 TRUSTED MESSENGERS TO AMPLIFY MESSAGES

In anticipating how they will counter electoral misinformation and disinformation, EMBs should consider who are the effective messengers they can call on for rapid amplification of crisis communications that can credibly reach the audiences that are at greatest risk:

- Who are the most effective messengers to reach supporters of different political parties?
- Who has credibility with groups that might be susceptible to violence or extremism in an instance where false information was rampant?
- Who can reach women, youth, or different religious communities?

Proactively identifying the right messengers can be a key preparatory step that allows an EMB to disseminate factual information most effectively under pressure. Preparing these networks to disseminate information to their communities in times of information crisis can be an essential way to ensure that an EMB's message is amplified by credible sources in periods when a flood of information might drown out authoritative actors speaking from their own, more limited, platforms. For example, ahead of 2020 local elections, the Brazilian Superior Electoral Court (TSE), partnered with one of the country's most popular soccer clubs to counter fake news.

In 2018, the office of the Prime Minister of Finland created an initiative to [work with social media influencers](#)



HIGHLIGHT



The Brazilian football club partnership disseminated content that used sports analogies, including this example referring to the VAR football verification system. In this instance, the content seeks to build confidence in election integrity and counter rumors of security flaws in Brazil's electronic voting machines by noting the widespread use of similar machines in other countries.

(<https://www.theguardian.com/world/2020/apr/01/finland-enlists-social-influencers-in-fight-against-covid-19>) to disseminate credible information in a crisis scenario. The network of 1,500 influencers that was established through that initiative was first activated to disseminate credible health information during the COVID-19 pandemic. Working with social media influencers enabled the government to reach audiences that are not consumers of mainstream media. For example, a video of an influential YouTube personality interviewing a government minister and health experts received more than 100,000 views within two days. A similar model could be employed by an EMB, for example, engaging social media influencers in advance of an election to sign a peace pledge that committed them to disseminating credible information and promoting peace on and directly after Election Day in a country where electoral violence is a concern.

ELECTION MANAGEMENT BODY APPROACHES TO COUNTERING DISINFORMATION

3. EMB CODES OF CONDUCT OR DECLARATIONS OF PRINCIPLE FOR THE ELECTORAL PERIOD (/TOPICS/EMBS/3- EMB-CODES-CONDUCT-OR- DECLARATIONS-PRINCIPLE-ELECTORAL- PERIOD)

While EMBs generally lack authority to sanction or deter the behavior of foreign disinformation actors, they may have a mandate to set standards and norms for domestic actors. Codes of conduct are a tool used by some EMBs to define how political parties, candidates, media or the electorate at large should behave during the electoral period. In recent years, some EMBs have moved to fill the normative and regulatory gap that exists around the use of social media in elections by creating codes of conduct, codes of ethics or declarations of principles (for the purposes of this subcategory, these are collectively referred to as codes of conduct, meaning documents outlining normative behaviors for the electoral period).

Codes of conduct (<https://aceproject.org/main/english/ei/eif01a1.htm>) can either be voluntary, non-binding agreements that result from a consensus among the parties, or they can be part of the legislative and regulatory framework that is binding and enforced. Codes of conduct for the

use of social media in elections include examples of both types. Voluntary, non-binding agreements tend to be shorter in length, committing signatories to broad principles. Those that have some weight of enforcement, of necessity, contain provisions that have greater specificity.

"[The Principles allow] us to say that our political parties agree on a set of rules and it is a first step in moving towards developed democracy where political opponents respect one another and demonstrate issue-based discussions. In the long term, having a culture of dialogue instead of negative campaigning and defamation of political candidates is the goal of this document." — IFES Interlocutor at the Central Election Commission of the Republic of Georgia

The guidebook section on [Norms and Standards \(/topics/norms/0-overview-norms\)](/topics/norms/0-overview-norms) discusses regional frameworks and other transnational examples of norm setting around disinformation. The guidebook section on [Legal and Regulatory \(/topics/legal/0-overview-legal-and-regulatory-responses\)](/topics/legal/0-overview-legal-and-regulatory-responses) approaches to countering disinformation discusses a larger array of legal approaches governing the use of social media in elections. This subsection is limited to codes of conduct that address disinformation (exclusively or in combination with other problematic electoral behaviors) and are created and promulgated by EMBs to govern the conduct of political parties, candidates and their supporters, or the media during elections.



HIGHLIGHT

The Election Commission of India created a “Voluntary Code of Ethics for the 2019 General Election” that was developed in consultation with representatives of Social Media Platforms to govern the behavior of these entities during 2019 elections. Additional details can be found in the subcategory addressing [EMB cooperation with social media and technology companies \(/topics/embs/7-emb-coordination-technology-and-social-media-companies\)](/topics/embs/7-emb-coordination-technology-and-social-media-companies).

3.1 AUDIENCE

EMB codes of conduct intended to limit disinformation can be directed toward various electoral stakeholders and can be limited to a specific election or exist as a standing document. The Central Election Commission of the Republic of Georgia, for example, narrowly tailored their counter-disinformation guidance in their [“Ethical Principles of Candidates of 28 October 2018 Presidential Elections \(https://cesko.ge/eng/list/show/115503-2018-tslis-28-oqtombris-saprezidento-](https://cesko.ge/eng/list/show/115503-2018-tslis-28-oqtombris-saprezidento-)

archevnebshi-monatsile-kandidatebis-etikis-printsipebis-prezentatsia)” to presidential candidates in the specified election. Panama’s Digital Ethical Pact (<https://www.tribunal-electoral.gob.pa/publicaciones/pacto-etico-digital/>) broadly addresses “users of digital media” in the context of elections. South Africa’s “Code of Conduct: Measures to Address Disinformation Intended to Cause Harm During the Election Period” (in draft form as of December 2020) is aimed at “every registered party and every candidate” with additional obligations under the code for how those parties and candidates must take appropriate recourse against any member, representative or supporter who behaves in violation of the code. Nepal’s “Code of Conduct to be followed by Mass Media, Non-Governmental Organizations and Observers (<https://www.election.gov.np/election/en/electoral-code-of-conducts1.html>)”¹ has chapters addressing different audiences.

Internal codes of conduct that political parties voluntarily adopt to govern the behavior of their candidates and members are discussed in the guidebook section on Political Parties (<https://counteringdisinformation.org/node/42/>). .

3.2 DEVELOPMENT PROCESS

Particularly in the case of codes of conduct that rely on the voluntary commitment of signatories, a consultative development process can increase the legitimacy of the document. In their 2015 guide (<https://www.idea.int/sites/default/files/publications/guidelines-for-the-development-of-a-social-media-code-of-conduct-for-elections.pdf>) on developing social media codes of conduct, International IDEA recommends that EMBs “engage in a consultation process with a broad range of electoral stakeholders, especially journalists, bloggers, government agencies, and political commentators, that begins in the pre-electoral phase of an electoral cycle.” Consultations with civil society actors who represent different marginalized groups is also encouraged.

In Indonesia, Bawaslu conducted a highly consultative process in the development of their declaration (<http://www.bawaslu.go.id/id/berita/100-stakeholder-deklarasi-tolak-politik-uang-dan-adu-domba-di-pilkada-2018-dan-pemilu-2019>) to “Reject and Counter Vote Buying, Insults, Incitements, and Divisive Conflict in the 2018 and Pilkada and 2019 General Election.” The pledge was signed by 102 participating organizations after a 3-day consultative event that included CSOs, universities, religious organizations, and youth groups.² Signatories committed to a seven-point declaration rejecting intimidation and disinformation. This consultative process created a network of known and trusted actors that Bawaslu continued to work with on issues of disinformation and incitement throughout the 2018 and 2019 electoral periods. In this instance, the process of creating the declaration and the network of actors that came out of it was of equal if not greater value than the substance of the code itself. Bawaslu’s coordinated, multi-stakeholder responses to disinformation are explored in more detail in the subsection on EMB Coordination with Civil Society (<https://staging.counteringdisinformation.org/topics/embs/8-emb-coordination-civil-society>).

3.3 COMMON ELEMENTS

Codes of conduct that address disinformation can take many different forms. In some countries, a commitment to refraining from sharing disinformation is included as part of a broader code of conduct that covers all forms of conduct during an electoral period. In others, a code to deter disinformation is created to stand on its own. Some codes are only a few hundred words in length; others are much longer. Despite these differences, there are several common elements that could be considered by other electoral authorities looking to develop their own standards:

<p>Definitions</p>	<p>Because the array of content that can be considered disinformation is relatively broad, it is necessary for electoral authorities to define the scope of violations that they view as falling under their authority. Particularly for codes of conduct that have some element of enforceability, the provision of clear and specific definitions is essential for enforcement.</p> <p>South Africa’s code is drawn narrowly to limit its application to the electoral period and ground it firmly in the broader legal and regulatory framework in South Africa. Disinformation is defined as “any false information that is published with the intention of causing public harm.” That reference to public harm is based in the 1998 Electoral Act, which defines “public harm” as “(a) disrupting or preventing elections; (b) creating hostility or fear in order to influence the conduct or outcome of an election; or (c) influencing the outcome or conduct of an election.” This narrowly drawn definition creates gates around the types of disinformation that fall under the responsibility of the EMB; the EMB’s code addresses false information, published with intent to threaten the integrity of the electoral process.</p>
<p>Commitment to Freedom of Expression</p>	<p>Any code of conduct designed to deter disinformation will place bounds on what speech is allowable in an electoral context. As outlined in international human rights declarations and many national constitutions, any limitations on freedom of speech must meet a strict degree of scrutiny. As such, multiple EMBs have opted to include explicit recognition of the commitment to freedom of expression in the text of the code itself.</p> <p>South Africa’s code, for example, includes an affirmation that efforts to curb disinformation must “tak[e] into consideration the right to freedom of expression” contained in the national Constitution.³ The introductory language to Panama’s Digital Ethical Pact outlines the challenges of disinformation and social media while noting that “it is important to remember that freedom of expression and respect for the civil and political rights that have been so difficult to achieve in a democracy, are and should continue to be, the guide for us to have a better Panama in the future.”⁴</p>

<p>Ban deliberate sharing of fake news</p>	<p>A core element across codes of conduct intended to limit disinformation is a provision exhorting signatory parties to refrain from knowingly sharing false information. This is drawn more or less narrowly and is framed differently in each code. The Georgian Ethical Principles include broad guidance to “abstain from dissemination of false information with prior knowledge,”⁵ but provide no additional details. Panama’s Digital Ethical Pact includes a call for signatories to be vigilant before the appearance of ‘fake news’ or false information that may endanger the electoral process, and imputes a proactive responsibility for signatories to seek reliable sources of information before sharing messages that may be false.⁶</p> <p>This prohibition against the intentional sharing of false information may have precedent in broader national electoral law and general codes of conduct, and may extend existing principles that cover traditional media or campaigning to the realm of social media more specifically. In South Africa, the (draft) disinformation code of conduct is meant “to give effect to the prohibition against intentionally false statements contained in section 89(2) of the Electoral Act [73 of 1998].” Nepal’s code, which covers all aspects of the electoral period, calls on the mass media “not to publish, broadcast or disseminate the baseless information in favor of or against [a] candidate or political party on electronically used social networks such as S.M.S. [sic], Facebook, Twitter, and Viber.”⁷</p>
<p>Restricting deceptive online behaviors used to promote campaign content</p>	<p>In addition to guidance or limitations on the type or quality of <i>content</i> that signatories can use during campaign periods, codes of conduct can also provide restrictions related to what online <i>behaviors</i> are outside the bounds of ethical campaigning. This most often takes the form of exhortations to refrain from using specific techniques of artificial or manufactured amplification in ways the EMB perceives to be unethical or deceptive.</p> <p>Panama’s Digital Ethical Pact, for example, instructs signatories to refrain from using false accounts and bots to misinform or promote electoral propaganda.⁸ Provisions of this nature must strike a difficult balance given that the disinformation tactics of malign actors continue to evolve. Too narrowly defining the discouraged online behaviors leaves open the door to a range of other tactics that are being used; too broadly-defined measures have little meaning or deterrent effect. Tying these tools to their potential deceptive uses, as Panama’s Pact does, is an important approach to strike that balance. A blanket ban on tools like bots would likely be overly onerous and prevent their legitimate use as, for example, part of an effort that provides information to voters on how to cast their ballot.</p>

Prohibitions
against
incitement
to violence
and hate
speech

In addition to discouraging the dissemination of false information, codes of conduct might also establish the expectation that candidates, parties or other signatories will refrain from incitement to violence or hate speech in campaigning.

Panama's Digital Ethical Pact instructs digital media users to avoid "dirty campaigns" that "offend human dignity through the use of insults, incursions into privacy, discrimination" or "promote violence and lack of tolerance."⁹ Georgia's Ethical Principles instructed the presidential candidates to "refuse to use any hate speech, or statements that involve xenophobia or intimidation." South Africa's code does not explicitly prohibit hate speech, but its definition of "public harm" includes content that "create[s] hostility or fear in order to influence the conduct or outcome of an election."¹⁰

Some codes of conduct also prohibit hate speech based on particular identity categories, including gender, and specifically prohibit violence against women in politics. Codes of conduct must include specific reference to gender-related hate speech and online violence and harassment against women in politics so that actors are held accountable for these specific acts. For example, Guyana's 2017 Code of Conduct for Media (<https://gecom.org.gy/archived/pdf/MEDIA%20CODE%20OF%20CONDUCT.pdf>) – developed through the election commission's engagement with leading media representatives – enjoined the media "to refrain from ridiculing, stigmatising or demonising people on the basis of gender, race, class, ethnicity, language, sexual orientation and physical or mental ability" in their coverage of campaigns and elections.¹¹

<p>Application of a social media ban to the campaign period</p>	<p>It is also possible to use a code of conduct as an opportunity to set standards for the behavior of signatories during the defined campaign period, which may include limitations on social media use during a silence or blackout period directly before Election Day. Panama’s Digital Ethical Pact requires signatories to “collaborate with the Electoral Tribunal so that the electoral ban is respected and electoral campaigning is only carried out during the allowed period 45 days before the internal elections of the political parties and 60 days before the general election.”¹² Nepal’s code stipulates that during the electoral silence period votes cannot be solicited through campaigning via social media or other electronic means.¹³ As discussed in the legal and regulatory (https://staging.counterinformation.org/topics/embs/4-social-media-monitoring-legal-and-regulatory-compliance) section of this guidebook the specifics about what types of content are restricted outside of the campaign period should be clearly defined. For example, authorities may choose to disallow paid advertising, while allowing organic posts on the personal accounts of candidates and parties.</p>
<p>Proactive obligation to share correct information</p>	<p>Codes of conduct may require signatories to not only refrain from sharing false information, but to actively work to correct false and problematic narratives that do circulate. South Africa’s draft code obligates parties and candidates to address disinformation, “including by working in consultation with the Commission to correct any disinformation and remedy any public harm caused by a statement made by one of their candidates, office-bearers, representatives, members or supporters....”¹⁴ While not yet observed in practice, including a proactive responsibility for parties and candidates to work with the election commission to counter false or problematic electoral narratives does provide the Commission with an additional avenue for disseminating corrections, counter-narratives or voter information messages as part of a crisis communication strategy. South Africa’s code also requires signatories to publicize the code and educate voters about it.¹⁵</p>

3.4 ENFORCEMENT

Codes of conduct, as noted above, can be voluntary and nonbinding agreements or they can operate in conjunction with the legal and regulatory framework, allowing some degree of enforcement. Both voluntary and enforceable codes establish normative standards for signatories of the document. For voluntary codes, establishing norms through the public commitment of candidates, political parties and other relevant electoral stakeholders might be the sole purpose of the code.

EMBs have varying degrees of legal authority and capacity to enforce codes of conduct. In the Georgian case, the decision to adopt a declaration of principles rather than a code of ethics was done, in part, out of a recognition that the CEC lacked an existing mechanism for implementation or enforcement.¹⁶ In the case of South Africa, the EMB's enforcement mandate predates the code on disinformation, as they also have enforcement capabilities in regard to the general Electoral Code of Conduct and in the broader legal framework. The South African code defines the boundaries of the EMB's enforcement capacity, noting, for example, that if the EMB considers any content that comes to them as a result of the code of conduct to be a violation of existing criminal laws, then it will be duly referred to the appropriate law enforcement agency.¹⁷ Similarly, the commission stipulates that it will refer complaints against members of the media to existing bodies that have oversight of the press.¹⁸

Even when codes are situated in a clear legal framework, they are less weighty than other types of legal or regulatory deterrents. Vice-Chairperson of the South African IEC, Janet Love, characterized the IEC's enforcement of the Digital Disinformation Code as "measured" rather than "aggressive."

"We can't pretend to have a bazooka when in reality we have a firm stick." -- Vice-Chairperson Janet Love, Electoral Commission of South Africa

Though codes of conduct carry less legal weight, they do provide a flexibility that may be very attractive to EMBs. An enforceable code of conduct may be more easily and expeditiously adopted and led by the EMB, in comparison to a regulatory or legislative reform process. An enforceable code can provide EMBs with a "firm stick" by which they can strongly encourage compliance without resorting to lengthy legal proceedings that may drag on too long to allow for timely remedy. Codes of conduct can also side step serious harms that could stem from using revisions to the criminal code as an alternate approach. A further discussion of the potential harms of criminalizing disinformation are discussed in the guidebook section on [Legal and Regulatory](#) ([/topics/legal/1-definitions](#)) approaches to countering disinformation.

ELECTION MANAGEMENT BODY
APPROACHES TO COUNTERING

DISINFORMATION

4. SOCIAL MEDIA MONITORING FOR LEGAL AND REGULATORY COMPLIANCE (/TOPICS/EMBS/4-SOCIAL-MEDIA-MONITORING-LEGAL-AND-REGULATORY-COMPLIANCE)

A limited number of EMBS have a mandate to monitor the social media use of candidates, parties, media outlets or other designated electoral stakeholders to ensure compliance with the legal and regulatory framework. Monitoring might seek to enforce legal limits on campaign spending on political advertising on social media or on campaigning outside of a designated campaign period, or to enforce restrictions on content that has been deemed illegal in the context of an election. For many EMBS, this responsibility is not part of their legal mandate. In these instances, a mandate to monitor and enforce may rest with another entity, such as a media or political finance oversight body or anti-corruption agency, and the guidance outlined in this subcategory would be applicable to their work. Developing a means to monitor social media for compliance must go hand-in-hand with the development of [legal and regulatory frameworks \(/topics/legal/0-overview-legal-and-regulatory-responses\)](/topics/legal/0-overview-legal-and-regulatory-responses) that govern the use of social media during campaigns and elections. Without establishing a capacity to monitor, audit or otherwise effectively provide oversight, laws and regulation governing the use of social media during elections are unenforceable.

In truth, developing effective mechanisms to monitor electoral stakeholders' online conduct is a challenge without ready solutions for many oversight bodies. Efforts to conduct effective monitoring are often highly dependent on the transparency tools made available by social media platforms. Facebook is ahead of other platforms in rolling out political advertising transparency tools and expanding them to more countries, but many countries still lack access and local users have criticized the Facebook Ad Library for not being comprehensive. Google's political advertising transparency tools are only available in the EU and a handful of other consolidated democracies, with no observable efforts in 2020 to expand the availability of these tools to more countries. Other platforms offer even more limited tools for political advertising transparency.

While a range of commercial tools do exist for aggregating social media content to aid in the analysis of the online messages and conduct of political actors, the lack of customization of these tools for use by oversight bodies remains a challenge. Commercial tools are also often costly. Anecdotally, multiple oversight bodies that are starting new efforts to monitor social media during elections have shared that at present their approach is largely manual, consisting of staff members visiting the individual pages and accounts of political actors or other electoral stakeholders to analyze the content that has been posted.

Some EMBs with an oversight mandate are, however, innovating and expanding their ability to monitor social media for legal and regulatory compliance. Bawaslu, for example, monitored the official social media accounts of political candidates during 2019 Indonesian elections, though they acknowledged the limitations of this effort, observing that candidates keep their official pages free from controversy, while any misleading or divisive content would be disseminated and amplified through social media accounts not officially associated with the campaign. This effort was part of a larger approach to monitoring social media in collaboration with the Ministry of Information and Communication and the Cyber Crime Police Unit, which included efforts to detect deceptive coordinated campaigns on social media that might have links to candidates or political parties.

The efforts of the High Independent Election Authority (HIEA) of Tunisia to establish a capacity to monitor social media during the electoral period is illustrative of some of the approaches and challenges that such an effort may employ or encounter. Tunisia's legal framework does not have explicit provisions governing the use of social media during electoral campaigns. However, during the 2019 electoral cycle, the election commission decided to monitor online content and social media to ensure that parties and candidates were respecting the principles and rules of the campaign. This work was undertaken as an extension of the work being done by the HIEA's Media Monitoring Unit (MMU), which looks at electronic and print media during electoral periods. While the MMU was able to surface insights into the use of social media during the election, it also ran into a challenge common to social media monitoring efforts – defining the boundaries of which accounts are subject to monitoring. In Tunisia, as in other countries, the vast majority of offenses were observed to come from undeclared pages and accounts rather than the official accounts of the candidates. This creates challenges as in most cases there is insufficient evidence to definitively attribute these violations to the candidate or campaign benefiting from the problematic content.

4.1 DEFINING A MONITORING APPROACH.

Does the EMB have a legal mandate to monitor social media?

Prior to launching an EMB-led social media monitoring effort, the legal and regulatory framework must be consulted to ascertain the following:

- Does the EMB have a legal mandate to undertake monitoring activities?
- If not, does this mandate lie with a different government entity that would



HIGHLIGHT

prevent the EMB from conducting their own monitoring efforts?

- What legal and regulatory guidance exists, if any, for the use of social media during election periods?
- If there are no specific provisions related to the use of social media, are there general principles for the conduct of candidates, parties or other electoral stakeholders during the campaign that can be reasonably applied or extended to social media?

What is the goal of the monitoring effort?

After consulting the legal and regulatory framework, an EMB must establish an objective for their monitoring effort. For example, is the objective:

- To detect political advertising on social media that takes place outside of the designated campaign period?
- To identify instances in which online behaviors violate the legal framework governing the abuse of state resources?
- To monitor the content posted by candidates and parties to ensure compliance with any legal guidance on refraining from hate speech directed at women or other marginalized groups, incitement to violence, disinformation about the election, or other prohibited messages?
- To verify that reported spending on social media political advertising is accurate?

If there is little or no current legal or regulatory guidance on the use of social media in elections, is the effort:

- To gather information and evidence to inform future law reform conversations or the development of a code of conduct?

WHAT IS MEANT BY "SOCIAL MEDIA MONITORING"?

An increasing number of EMBs are identifying the ability to monitor social media as a skill that would aid them in fulfilling a counter-disinformation mandate. However, there are two different functions that are commonly implied by the phrase "social media monitoring."

The first function is monitoring the social media use of candidates, parties, media outlets or other designated electoral stakeholders for the purposes of **ensuring compliance with legal and regulatory guidance**. This function is intimately linked to the detection of violations and is necessary for enforcement of the legal and regulatory framework as it applies to social media.

The second function that is often implied by the phrase "social media monitoring" might more accurately be described as "social listening." Rather than monitoring the behavior of certain actors, social listening is an attempt to **distill meaning from the broad universe of conversations that are happening on social media** and other online sources in order to inform appropriate action.

These two functions are explored under separate subcategories: (4) *Social Media Monitoring for Legal and Regulatory Compliance* and (5) *Social Listening and Incident Response*.

- To raise public awareness about problematic content and behaviors that parties and candidates are engaging in on social media, including online harassment and violence against women and other marginalized groups?

What is the time-period for social media monitoring?

Based on the goal that is identified, the EMB should define how far in advance of elections social media monitoring efforts will begin, and whether they will extend to any part of the post-electoral period.

Will monitoring be an internal operation or will the EMB coordinate with other entities?

An EMB will need to ascertain whether it has sufficient capacity to conduct a media monitoring effort independently:

- Does the EMB have the human capacity and financial resources to conduct their own monitoring effort?
- Are there other state agencies or oversight bodies monitoring social media during elections that should be consulted or partnered with before an EMB launches their own effort?
- Are there any restrictions or prohibitions that would limit the EMB's ability to procure outside services from the private sector to augment the EMB's capacity?
- If the objective of the monitoring effort is to gather information and evidence to inform future law reform conversations or to understand how certain marginalized groups are targeted by disinformation, is there a role for credible civil society actors focused on advocating for legal reform or representing marginalized groups to provide the EMB with additional information and analysis?

What kinds of social media ad transparency tools are available in-country?

Understanding what is feasible for an EMB to do is in part contingent on the tools that technology and social media companies have made available in their country:

- Will the Facebook Ad Library¹ be enforcing disclosure for political and issue advertising in the relevant country?
- Will a Google Transparency Report² covering political advertising be available for the country?
- Do any other widely-used social media platforms offer transparency reports or features of any kind?
- If yes to any of the above, is the EMB equipped to use these tools to execute their mandate or is training necessary?
- Does the EMB have the authority to make legally binding requests of the social media platforms for information as part of an enforcement effort?

Further discussion of the definitional considerations necessary for establishing a social media monitoring approach is found in the guidebook section on [Legal and Regulatory responses](https://counteringdisinformation.org/node/2704/) (<https://counteringdisinformation.org/node/2704/>) to disinformation.

4.2 TYING SOCIAL MEDIA MONITORING TO A RESPONSE

Based on the identified goal of the social media monitoring effort, the EMB should identify how they will make use of the insights they gain through their monitoring efforts.

- If a legal and regulatory framework defining violations is in place, the EMB should identify how cases will be referred to appropriate enforcement agencies for further investigation and possible sanctions, if they do not have the ability to issue sanctions themselves.
- If the identified goal is to inform future regulation or the development of a code of conduct, a plan should be made for how content or behaviors that might constitute a violation under a revised legal framework is documented and explained in a way that would be compelling for the necessary audience of regulators or lawmakers.
- If the goal is to deter bad behavior by raising public awareness about the questionable or illegal conduct of parties and candidates, the public relations or communications department of the EMB should be involved in developing a plan for communicating findings to the general public.

Responses must take into account gender considerations and, in particular, should ensure that violations targeting marginalized groups or exploiting stereotypes about marginalized groups are specifically addressed so that these groups are not further marginalized by responses that are blind to their concerns and experiences.

ELECTION MANAGEMENT BODY APPROACHES TO COUNTERING DISINFORMATION

5. SOCIAL LISTENING TO UNDERSTAND AND RESPOND TO DISINFORMATION THREATS (/TOPICS/EMBS/5-SOCIAL-LISTENING-UNDERSTAND-AND-RESPOND-DISINFORMATION-THREATS)

Rather than monitoring the behavior of certain actors, social listening is an attempt to gain insights into the sentiment, misperceptions, or dominant narratives circulating on social media and other online forums in order to inform appropriate action. An EMB may wish to set up social listening to inform a rapid incident response system or to inform strategic and communication

planning. Gaining insight into *what* narratives are circulating and gaining popularity in online spaces can provide EMBs with insights into *how* to effectively counter narratives that threaten election integrity.

If electoral authorities wish to monitor political parties or other electoral stakeholders for compliance with the legal and regulatory framework, please refer to the prior subsection on [Social Media Monitoring for Legal and Regulatory Compliance](/topics/embs/4-social-media-monitoring-legal-and-regulatory-compliance). (/topics/embs/4-social-media-monitoring-legal-and-regulatory-compliance)

5.1 UNDERSTAND EMB CAPACITY AND PURPOSE

Setting up a social listening and response capacity is not a one-size-fits-all effort. Some EMBs have the staff capacity, recognized mandate, and financial resources to set up comprehensive efforts. For other EMBs, the barriers to entry to establishing social listening capacity may seem (or be) insurmountable and may divert attention from more essential activities. If donors and international assistance providers are assisting an EMB to establish or strengthen a social listening capacity, it is essential to tailor the monitoring effort to fit the EMB's needs and capacity.

EMBs will have different purposes for setting up social listening capacity. This subsection focuses primarily on EMBs that wish to build a real-time monitoring effort that allows them to identify and respond to disinformation or other problematic content swiftly. Other EMBs may wish to use social listening earlier in the electoral cycle to inform communication strategies. This proactive strategy is briefly discussed in this subcategory for ease of comparison with other reactive applications of social listening. These efforts are not mutually exclusive and an EMB may choose to pursue both.

5.2 SOCIAL LISTENING TO INFORM RAPID INCIDENT RESPONSE

The National Electoral Institute (INE) of Mexico's social listening and incident response efforts illustrate what a fully-staffed and resourced social listening effort can look like. INE designed and deployed "Project Certeza" in the days prior to and on Election Day in 2018, and also implemented the same system for 2019 elections. Project Certeza's purpose was to "identify and deal with false information disseminated, particularly through social networks but also through any other media, that could produce uncertainty or distrust in the citizenry about the electoral authority's responsibilities as the election is happening."¹ This effort included a technological monitoring system developed by INE, which screened millions of pieces of social media content and other sources for potentially problematic words and phrases associated with elections. That flagged content was then referred to human moderators for verification and determination on whether the content required action. In addition to this remote monitoring, INE hired a network of temporary field operators to gather real-world information and document first-hand evidence that

could be used to refute false and inaccurate claims.² Evidence and analysis from the remote monitoring team and field teams were then shared with INE's social outreach division, where specifically-tailored refutations or voter information content was shared via social networks and with media outlets. The team working on Project Certeza included senior officials from eight different divisions at INE, which meant that immediate decisions could be made on appropriate responses.³ An effort as comprehensive as Mexico's will be beyond the reach of most EMBs. However, elements may still be illustrative to other EMBs designing their own social listening efforts.

As an alternative to such an approach, election authorities might consider interventions that help voters encounter reliable information when they seek more details about a piece of disinformation that they have encountered. Election authorities in the U.S. State of Colorado monitored social media to identify trending misinformation and disinformation about the U.S. 2020 elections and then [purchased Google ads](https://www.nytimes.com/2020/10/20/us/politics/election-colorado-misinformation.html?referringSource=articleShare) (<https://www.nytimes.com/2020/10/20/us/politics/election-colorado-misinformation.html?referringSource=articleShare>) tied to relevant search terms. This was an attempt to ensure that information seekers using the search engine to look up the disinformation they encountered were directed to credible sources, rather than surfacing search results that further fed conspiracy. Placing Google ads to ensure credible results appear at the top of a search page can be one approach to combat disinformation that emerges through "[data voids](https://datasociety.net/wp-content/uploads/2019/11/Data-Voids-2.0-Final.pdf) (<https://datasociety.net/wp-content/uploads/2019/11/Data-Voids-2.0-Final.pdf>)," which can occur when obscure search queries have few results associated with them, making it easier for disinformation actors to optimize their content in ways that ensure information seekers encounter content that confirms rather than rebuts disinformation.

Another prospective area for social listening that might be better suited to EMBs that lack internal capacity to set up an independent effort is partnering with a technical assistance provider, working with civil society or contracting a credible private entity that specializes in social listening to set up an early warning system of alerts that could be monitored by EMB staff. Alerts could be built around key phrases, such as the name of the EMB, that would be triggered when social media content containing those phrases starts to go viral. The alerts could be designed based on high likelihood, high impact scenario planning that might be included as part of the development of a [crisis communication strategy](/topics/embs/2-crisis-communication-planning-disinformation-threats) (</topics/embs/2-crisis-communication-planning-disinformation-threats>). For example, an EMB might determine that voter registration in a particular region or the integrity of overseas voting are topics at high risk of being the subject of damaging mis- or disinformation. By anticipating these scenarios, the EMB could tailor alerts that would flag potentially problematic content as it starts to gain popularity.

This approach would be considerably less comprehensive than a well-staffed internal monitoring effort, but for EMBs that lack more robust options, limited solutions may still have value. This research has not surfaced any examples of EMBs using this strategy currently, but a network of civil society actors in Slovakia, including media monitoring and elections CSO Memo98, used a similar model to set up a series of alerts for the Slovak Health Ministry to notify them of trending

misinformation and disinformation about COVID-19. The ability of their counterparts at the Ministry to use the alerts in actionable ways was limited, suggesting that any initiative of this nature must be carefully planned to meet the needs and capacity of the EMB.

Existing methodologies for detecting online violence against women in elections could be adapted to assist EMBs in understanding the ways in which gendered messages are contributing to distortions of the information environment around elections and to craft more impactful responses based on these insights. CEPPS has used AI-informed social listening to monitor online violence against women in elections (<https://www.ifes.org/publications/violence-against-women-elections-online-social-media-analysis-tool>), and findings and lessons learned from this work could be used to inform disinformation programming. Lessons learned from this work confirm that automated data mining techniques only go so far in distinguishing problematic content, and that the combination of automated techniques and human coders is essential to having accurate insights.

5.3 SOCIAL LISTENING TO INFORM STRATEGIC AND CRISIS COMMUNICATION PLANNING:

Social listening can be integrated into the development of communication strategies, providing insights into how electoral processes, the information environment and the EMB are perceived among different demographic groups. This understanding can in turn help an EMB craft evidence-based communication strategies to reach different audiences.

To inform its strategic and crisis communication planning, the Independent Electoral and Boundaries Commission (IEBC) of Kenya worked with a social listening firm to receive an overview of the social and digital media landscape in Kenya prior to 2017 elections. Insights gained through social listening are made more valuable through further analysis; the outside firm combined insights from their social media analysis with findings from a series of focus group discussions that explored awareness and perception of various digital platforms, as well as understanding how different sources of information were used by voters and the motivations behind sharing “fake news,” misinformation, and hate speech. Focus group participants also shared perceptions of the IEBC and provided feedback on the persuasiveness of sample messaging strategies. Engaging outside experts to conduct this analysis can supplement the EMB's capacity.

5.4 DEFINING A MONITORING APPROACH

Given the variations in need and capacity, each monitoring approach must be calibrated to suit



HIGHLIGHT

the institution that uses it.

What is the goal of the social listening approach?

Examples of insights EMB can gain through social listening include:

How the EMB is being talked about on social media.

Given that one goal of anti-democratic influence operations is to undermine trust in electoral processes and institutions, social listening can help an EMB engage in some degree of “reputation management.” Social listening can give insights into where EMB performance may be seen as lacking, can help explain any accusations directed toward the EMB, or can help EMBs understand where a lack of transparency in their operations might generate distrust.

Whether false or problematic narratives about elections are gaining traction on social media.

As part of an Election Day incident response plan, an EMB can monitor social media for allegations of malpractice, fraud, or violence in certain regions or at particular polling stations that need to be corrected or acknowledged. They can also use this information to determine how to distribute resources or support to districts or polling stations that are experiencing difficulties.

Whether misinformation or disinformation is circulated that might suppress voter turnout or otherwise impact the integrity of the election.

Based on their crisis communication planning, EMBs can determine when and how they will respond to voter interference messages that they might detect circulating on social media. If social listening reveals ways in which certain populations are being targeted as subjects or consumers of disinformation, for example, an EMB could use that information to focus counter messages toward impacted populations.

What is the time period for social listening?

WHAT IS MEANT BY "SOCIAL MEDIA MONITORING"?

An increasing number of EMBs are identifying the ability to monitor social media as a skill that would aid them in fulfilling a counter-disinformation mandate. However, there are two different functions that are commonly implied by the phrase “social media monitoring”:

- Monitoring the social media use of candidates, parties, media outlets, or other designated electoral stakeholders to **ensure compliance with legal and regulatory guidance.**
- Engaging in “social listening”, or the attempt to **distill meaning from the broad universe of conversations that are happening on social media** and other online sources to inform appropriate action.

Full descriptions of these functions can be found in the [prior subsection](#).

(<https://counteringdisinformation.org/topics/embs/social-media-monitoring-legal-and-regulatory-compliance>).

An EMB must determine how far in advance of elections social listening efforts will begin. Depending on resources and the goals of the social listening exercise, EMBs may choose to monitor only a narrow window of time around Election Day, or they may choose to monitor the entirety of the campaign period. EMBs using social listening for rapid incident response should also plan to continue efforts through the immediate post electoral period, when false and misleading information with the potential to incite violence or delegitimize results may be at its highest.

For EMBs using social listening to inform their strategic or crisis communication plans, EMBs must strike a balance between completing this work far enough in advance to have strategies in place in time for the election, but not so far in advance that voters' opinions about the information environment are outdated by Election Day.

Will the social listening effort be an internal operation or will the EMB partner with other entities?

An EMB will need to ascertain whether it has sufficient capacity to conduct a social listening effort independently:

- Does the EMB have the capacity and resources to conduct their own social listening effort?
- Are there other state agencies, civil society organizations or academics conducting similar work that might be able to partner with the EMB to do this work?
- Are there any restrictions or prohibitions that would limit the EMB's ability to procure outside services from the private sector to augment the EMB's capacity?

Which tools will the EMB use to monitor social media platforms or other online sources?

If an EMB does not have the capacity to develop their own system, as Mexico's INE did, a range of social listening tools are available. Those that are most comprehensive are available through paid subscription. Many of these tools and possible applications are discussed in NDI's publication [Data Analytics for Social Media Monitoring \(https://www.ndi.org/publications/data-analytics-social-media-monitoring?eType=EmailBlastContent&eld=06cd3edc-c970-4db3-afe8-294e972a4069\)](https://www.ndi.org/publications/data-analytics-social-media-monitoring?eType=EmailBlastContent&eld=06cd3edc-c970-4db3-afe8-294e972a4069).

5.5 TYING SOCIAL LISTENING TO ACTION

The purpose of engaging in social listening is to inform more effective responses from the EMB. To that end, social listening for the purpose of rapid incident response should be closely aligned with an EMB's crisis communication planning. Based on scenario planning done during crisis communication planning, the EMB should map out what their process will be for responding to any problematic or misleading content that they identify through social listening. There should be clear lines of internal communication for verifying suspect content. This process may include receiving rapid input from regional election commissions or individual polling stations. Communication channels, including traditional media actors or identified trusted messengers, should also be established in advance.

Additionally, social listening may surface cases that may be referred to another government entity. For credible reports of activities in violation of the criminal code, the EMB should be prepared to refer reports to the appropriate actor. For example, INE's social listening efforts in the 2019 Mexican elections surfaced three credible reports of vote buying that were referred to the Special Attorney on Electoral Crimes.⁴

ELECTION MANAGEMENT BODY APPROACHES TO COUNTERING DISINFORMATION

6. DISINFORMATION COMPLAINTS REFERRAL AND ADJUDICATION PROCESS (/TOPICS/EMBS/6- DISINFORMATION-COMPLAINTS- REFERRAL-AND-ADJUDICATION- PROCESS)

Given controversy and lack of consensus over the standards by which social media platforms determine what content is allowable on their platforms, increasing national sovereignty over what content is allowable is of interest in many countries. The IEC in South Africa and Bawaslu in Indonesia have adopted disinformation complaints and adjudication processes to increase national decision-making power over the removal of certain types of content during electoral periods from social media platforms.

"If you submit a complaint to Twitter or Google – your complaint [is] adjudicated according to the terms of a private company. If you don't like the outcome, there is nothing you can do about it, there is no transparency. It means a foreign entity is determining things of national importance." William Bird – Director of Media Monitoring Africa

Rather than flagging content via native reporting functions within the social media platform and leaving it up to the company's discretion and community standards to remove or downrank that content, Bawaslu and the IEC developed processes that allowed them to issue a decision with the force of law to compel the platforms to remove content. For this approach to be given

consideration, EMBs must be independent and credible institutions. If an EMB is not sufficiently independent of political pressures, a process such as this could be easily abused for political advantage.

South Africa: Ahead of May 2019 Elections, the IEC worked with civil society to establish a complaints referral and adjudication process. An online portal was launched that allowed the public to lodge complaints about specific pieces of content. Complaints were received by the IEC's Directorate of Electoral Offences, which worked with a Digital Disinformation Commission (DDC) composed of outside media, legal and technology experts to assess the complaints and make recommendations to the Commission for action. Decisions were communicated to the public by the IEC through regular reports to the media and the status of complaints as they made their way through the process was publicly tracked on the IEC's website.

Indonesia: Ahead of April 2019 Elections, Bawaslu established a complaints referral and adjudication process. In addition to receiving complaints directly from the public, Bawaslu also received a weekly compilation of complaints that had been received by the Ministry of Communication and Information Technology. A part-time task force within Bawaslu would then assess and categorize the content to determine if it was in violation of national standards. Content that was determined to be in violation was sent to the social media platforms via an accelerated channel for review and removal. Of the 3,500 complaints received by Bawaslu, 174 were determined to be within their jurisdiction (related to elections) and processed for further review by the platforms.

Setting up a complaints referral and adjudication process is a labor- and resource-intensive undertaking. If an EMB is contemplating this approach, there are several factors that should be considered regarding whether they should set up a system, and if they do, how to design an effective system.

6.1 THROUGH WHAT AVENUES IS CONTENT REPORTED?

As with any complaints adjudication process, the designers of the system must consider who has standing to bring a complaint for consideration. Should anyone have the ability to flag a piece of content for review by the EMB? Only political parties and candidates? Should the EMB aggregate complaints received by other government agencies or bodies?

The South African system allows that "the Commission shall receive complaints of disinformation during the election period from *any person*."¹ This is operationalized through the Real411 system (<https://www.real411.org/>), which includes a web portal where any member of the public, regardless of whether they are eligible voters, can flag content for review. The portal now receives complaints year-round, not only during the electoral period, and is maintained by the civil society

group Media Monitoring Africa, which organizes the three-person review teams of outside experts that make up the Digital Disinformation Commission (DDC). During elections, the DDC makes recommendations to the IEC's Directorate of Electoral Offenses on actions to be considered.

Systems that are open to public reporting from any member of the public provide opportunities for brigading, in which actors wishing to overwhelm or discredit the system could flood the reporting channel with disingenuous or inaccurate reports. A non-disinformation example of this took place ahead of 2020 Serbian elections. In this instance, a party that was boycotting the elections created a viral Facebook campaign encouraging supporters to submit claims via the EMB's election complaints process for the suspected purpose of overwhelming the EMB's dispute resolution capacity. Though the cases were dismissed, they reportedly caused administrative delays which weighed on the effectiveness of the complaints process. South Africa attempts to mitigate against this risk by requiring complainants to confidentially submit their names and email addresses along with their complaint.

Bawaslu, recognizing the ways in which overly formal reporting mechanisms can significantly slow collaboration, maintained informal communication channels with counterparts at the Ministry of Communication and Information Technology, the police, and the army via WhatsApp to refer and share intelligence about complaints in addition to receiving complaints directly from the public. On a weekly basis, the Ministry of Communication and Information Technology would collect and send the reports of content they had collected to Bawaslu for review, classification and a determination on further action. Interlocutors at Bawaslu estimate that they received an average of 300 to 400 reports per week.

6.2 WHAT ARE THE STANDARDS TO DETERMINE VIOLATING CONTENT

The definitions that a disinformation complaints referral and adjudication body uses to determine what content constitutes a violation that requires remedy or redress must be clearly and narrowly drawn and fit within the country's constitutional, legal and regulatory framework.

The screenshot shows the 'SUBMIT A COMPLAINT' page on the REAL 411 website. The page has a dark blue header with navigation links: HOME, ABOUT, SUBMIT A COMPLAINT, VIEW COMPLAINTS, MEDIA, and ELECTIONS. Below the header is a sub-header with the text: 'Help us stop the spread of disinformation, put a stop to hate speech and halt journalist harassment. You can view previous reports, report something you've seen and find out more about the commission by following these links.' The main heading is 'SUBMIT A COMPLAINT' in bold. Below it is the 'CONTACT INFORMATION' section, which includes a note: 'We'll use this information to contact you should we need more information, and to keep you updated on the status of your complaint. Your contact details are kept confidential.' The form fields are: 'YOUR NAME', 'YOUR EMAIL', and 'YOUR TELEPHONE' (with a dropdown for country code and a text input for the number). Below this is the 'COMPLAINT INFORMATION' section, which starts with a note: 'Please select from the complaint options below. Provide as much information as possible, the more information there is the more it will help us.' There are four radio button options: 'ARE YOU REPORTING MIS- OR DISINFORMATION?', 'ARE YOU REPORTING HATE SPEECH?', 'ARE YOU REPORTING INCITEMENT TO VIOLENCE?', and 'ARE YOU REPORTING JOURNALIST HARASSMENT?'. Each option has a small explanatory text below it. Below the radio buttons are two text input fields: 'WHERE DID YOU SEE THIS?' and 'WHAT LANGUAGE IS THIS IN?'. At the bottom of the form are two checkboxes: 'AGREE TO TERMS AND CONDITIONS' and a blue 'SUBMIT' button.

In South Africa, the complaints process is integrated into the draft Code of Conduct for Measure to Address Disinformation Intended to Cause Harm During the Election Period. The code itself draws clear definitions of what constitutes disinformation – specifically, intent to cause public harm, which includes disrupting or preventing elections or influencing the conduct or outcome of an election. As discussed in the subsection on [codes of conduct](https://counteringdisinformation.org/topics/embs/3-emb-codes-conduct-or-declarations-principle-electoral-period), (<https://counteringdisinformation.org/topics/embs/3-emb-codes-conduct-or-declarations-principle-electoral-period>) the code's definitions are firmly grounded in the South African Constitution and electoral legal framework. The same standards are used in each phase of the complaints process, by the DDC, which is external to the IEC, as well as by the Electoral Offenses Office and the Commissioners within the IEC.

Arriving at standardized definitions presents an opportunity for EMBs to engage in consultation and relationship building with potential allies in the fight against electoral disinformation. In Indonesia, Bawaslu created standardized definitions for unlawful content in electoral campaigns. Prior to 2018 local elections, existing laws outlined categories of prohibited content, such as hate speech, slander, and hoaxes, but these categories lacked clear definitions. To arrive at definitions, IFES supported Bawaslu in conducting a series of roundtable discussions engaging more than 40 stakeholders from government, civil society and religious organizations to discuss definitions for the types of content prohibited in electoral campaigns. This feedback was then taken into consideration in the formulation of Bawaslu's Regulation on Prohibited Electoral Campaign Content.² Consultation can be more narrowly drawn as well; in the run up to the launch of their complaints process, definitions in South Africa were discussed by a working group that included IEC members, media lawyers and members of the press.

Definitions are also likely to evolve over time as the complaints process is put to the test. In South Africa, initial discussions included whether hate speech and attacks on journalists should be covered by the complaints process. Though both were excluded from the definitions used during 2019 elections, Media Monitoring Africa and their partners developed definitions and reporting processes for these additional categories of complaints after the elections. Complaints on these additional topics are now able to be submitted via the complaints portal, and may be considered by the IEC for future elections.

It might also be useful to examine existing legal and regulatory frameworks around gender-based violence, violence against women, or gender equality that can be used to create definitions for online content that may violate these laws and regulations. Including definitions specific to violations that disproportionately affect women and other marginalized groups is key in making sure their concerns and experiences are addressed through this effort.

6.3 REMEDIES, SANCTIONS, AND ENFORCEMENT OF DECISIONS

An adjudication process should provide for a variety of remedies and sanctions that can be adapted to fit the violation that is identified. It may be desirable for a complaints adjudication process to have more remedies at its disposal than the referral of content to the platforms for

removal.

In both South Africa and Indonesia, the most common judgement regarding the referred complaints was to take no action - either because the content was not deemed to rise to the threshold of constituting public harm or because the content fell outside of the narrow focus on election-related content and therefore was outside the jurisdiction of the EMB.

The South African IEC has discretion to determine appropriate avenues for recourse. These include:³

- Determining that no action is necessary
- Engagement with the party or candidate that has committed the violation to urge compliance with the disinformation code of conduct, which stipulates that signatories must act to correct disinformation and remedy public harm in consultation with the IEC, including disinformation that originates with the signatories' representatives and supporters.
- Referral to the appropriate regulatory or industry body that has jurisdiction, including the Press Council of South Africa or the Independent Communications Authority of South Africa
- Referral to a relevant public body, such as the police, for further investigation or action
- Referral to the Electoral Court for appropriate penalty or sanction
- Use of IEC communication channels to correct disinformation and remedy public harm

The remedies envisioned through Bawaslu's process were narrower than those of their South African counterparts. Bawaslu had authority to observe social media during the campaign period, but not to take action against violators. The primary focus of their process was to elevate content for review and removal by the platforms. In instances where content was in violation of platform community standards, the content was removed or its distribution limited in accordance with platform policies. For Facebook, in instances where content was in violation of Indonesian law but did not violate Facebook's community standards, content was 'geoblocked' - meaning that the post was inaccessible from within Indonesia, but was still accessible outside of the country.

In addition to content removal or restriction, Bawaslu also used the content they collected to identify voter education and voter information themes to emphasize in their public messaging. They also referred cases to the criminal court system in instances where content violated the criminal code. Bawaslu reported that there were no instances of sanctions against political parties using the criminal code, though actions were taken against individuals. Notably, the highly-publicized "seven containers hoax" which alleged that cargo ships (<https://www.straitstimes.com/asia/se-asia/jakarta-probing-online-claims-about-containers-filled-with-ballots-for-president-jokowi>) full of pre-voted ballots had been sent to Jakarta, led to criminal charges (<https://www.benarnews.org/english/news/indonesian/ballot-hoax-04042019153456.html>) against the individuals that started and spread the hoax.

6.4 HOW WILL THE SYSTEM ACT EXPEDITIOUSLY?

The possible timeline for adjudication and action is a significant challenge for complaints referral and adjudication processes. Though some content identified as high priority was expeditiously addressed, the systems that Indonesia and South Africa developed and used during their respective elections had multiple stages that at times took weeks to clear in order to issue a decision on an individual piece of content. Given the volume of posts, quick iteration of messages and tactics, and the speed with which problematic content can go viral, a slow process for removing individual pieces of content is unlikely to have a measurable impact on the integrity of the information environment. By the time a piece of content has been in circulation for a day or two – much less a week or two – it is likely to have done the majority of its damage, and the churn of content will ensure that new narratives will have emerged to occupy public attention.

In instances where the remedy or sanction sought goes beyond content removal, as in the case of South Africa, a slower timeline may not reduce the effectiveness of the remedy. Media Monitoring Africa would at times pursue a dual track in which content would be referred to the IEC and to the platforms simultaneously: to the IEC for consideration of the array of remedies under their power to issue, and to the platforms for review of the content for expeditious removal. However, if the primary remedy sought by a complaints adjudication process is the removal of content from a social media platform, a multi-step complaints referral process may not be an efficient way to achieve it. Content removal may not be a goal that EMBs should be involved in at all.

6.5 HOW IS PUBLIC OUTREACH/PUBLIC AWARENESS RAISING BEING DONE?

The goals of a complaints referral and adjudication process should be twofold; the intent of such a system is to both remedy the harms of disinformation, as well as to build confidence among the electorate that authorities are effectively addressing the challenges of disinformation in ways that protect the integrity of the electoral process. Thinking through a communication strategy to publicize the efforts and successes of the process is a critical component to making the most of a complaints system. The very existence of the complaints system, if compellingly communicated to the public, can help to rebuild public perception of the trustworthiness of democratic processes and election results.

In the case of South Africa, a subsidiary benefit of running the complaints process was that the IEC could offer reassurances to the public after the election that the integrity of the election had not been undermined by coordinated malign actors seeking to distort outcomes or disrupt election processes. The referral mechanism in some ways also served as a crowdsourced media monitoring effort, and contributed to the conclusion that there was no evidence of foreign or state-linked influence operations that were operating at scale. The IEC concluded that there were instances of misinformation and disinformation, but no evidence of a coordinated disinformation campaign.

The complaints process developed by the IEC included planning for how to keep the public informed about the decisions that were made. As a way to build public awareness and interest in the complaints system, the IEC provided regular reports to the media that summarized the

complaints that were received and how they were handled. Though the complaints process was only active for a brief period before elections, the IEC's communication efforts helped build support for the complaints process, leading to calls for the system to continue even after the election.

6.6 PROVIDE ADEQUATE TIME TO DEVELOP AND REVIEW THE COMPLAINTS PROCESS

An effective complaints adjudication process is a complex endeavor to start and to gain institutional buy-in. Such a system may take time for implementers to learn to use. The ramp-up time for a system may be extensive, particularly if it involves consultative elements and involves developing common definitions. Any existing system should be reviewed in advance of each election to ensure it is suited to the evolving threat of electoral disinformation.

Plans and consultations for South Africa's Real411 System began in the fall of 2018 and the system was only able to begin operation in April 2019 ahead of May elections. Consultations on definitions of violating content in Indonesia, though at the time unconnected with the complaints referral system, took place a year in advance of 2019 elections.

ELECTION MANAGEMENT BODY APPROACHES TO COUNTERING DISINFORMATION

7. EMB COORDINATION WITH TECHNOLOGY AND SOCIAL MEDIA COMPANIES (/TOPICS/EMBS/7-EMB-COORDINATION-TECHNOLOGY-AND-SOCIAL-MEDIA-COMPANIES)

Coordination between EMBS and Technology and Social Media companies to enhance the dissemination of credible information or restrict the spread of problematic content during electoral periods.

Technology and social media companies – including but not limited to Facebook, Google, Twitter, TikTok, and their subsidiary companies including Instagram, WhatsApp, and YouTube – have a role to play in ensuring that elections take place in a credible information environment and that their

platforms are not used to undermine the integrity of elections. While companies must be held to account for harms that may stem from their platforms and services, progress towards alleviating these harms can be enhanced through direct engagement with these companies.

“Of course technology companies have more tools for how to regulate what happens on their platforms. There are things governments can’t do, but technology companies can – we need their help, but you need boldness from government to say so.” – Commissioner Fritz Edward Siregar, General Election Supervisory Agency of Indonesia (Bawaslu)

The market size of a country matters when it comes to how many resources social media companies are willing to invest and how available they are to assist and coordinate with EMBs. Before the 2019 elections, the Election Commission of India was able to convene representatives from top social media companies (<https://www.dw.com/en/india-fights-fake-news-on-social-media-ahead-of-election/a-48066548>) for a two-day brainstorming session on approaches to problematic social media content in elections, gaining a commitment from those companies to abide by a code of ethics. (<https://counteringdisinformation.org/topics/embs/3-emb-codes-conduct-or-declarations-principle-electoral-period>) Conversely, EMBs of smaller countries have reported difficulty getting company representatives to respond to their messages, even after establishing a point of contact within the company. There is significant variation in EMBs’ experiences working with social media and technology companies, as platforms may dedicate variable levels of support to specific countries based on factors including market size, geopolitical significance, potential for electoral violence, or international visibility.

“We aren’t naïve – these are profit-driven companies.” – Dr. Lorenzo Córdova Vianello, Councilor President of the National Electoral Institute of Mexico

There is also variation among social media platforms in terms of how willing they are to engage and how many resources they have put behind working with local election authorities. Indonesian electoral authorities, for example, reported that Facebook and YouTube had local representatives that made working with them easier, but that Twitter lacked the capability on the ground, making recurrent engagement more difficult. Based on conversations with more than two dozen EMBs globally, it appears that Facebook has invested more dedicated attention and resources than other platforms in establishing connections with election authorities in a wider range of countries.

The style and formality of agreements between tech companies and EMBs also varies from country to country. Ahead of 2018 Mexican elections, INE in many ways piloted what coordination with social media companies could look like, signing cooperative agreements with Facebook, Twitter and Google, for example.¹ The Brazilian TSE, building on INE's experience, signed formal agreements (<https://www.justicaeleitoral.jus.br/parcerias-digitais-eleicoes/>) with WhatsApp (https://www.justicaeleitoral.jus.br/parcerias-digitais-eleicoes/assets/arquivos/memorando_whatsapp.pdf), Facebook and Instagram (https://www.justicaeleitoral.jus.br/parcerias-digitais-eleicoes/assets/arquivos/memorando_facebook.pdf), Twitter (https://www.justicaeleitoral.jus.br/parcerias-digitais-eleicoes/assets/arquivos/memorando_twitter.pdf), Google (https://www.justicaeleitoral.jus.br/parcerias-digitais-eleicoes/assets/arquivos/memorando_google.pdf) and TikTok (https://www.justicaeleitoral.jus.br/parcerias-digitais-eleicoes/assets/arquivos/memorando_tiktok.pdf) ahead of 2020 elections. Brazilian electoral authorities pushed to include more concrete measures and actions to be adopted by the social media platforms, getting commitments from the platforms to use their features and architecture to react to malicious and inauthentic behavior, as well as to promote the dissemination of official information. The majority of arrangements, however, are less formal, and social media companies seem less willing to sign formal MoUs in some countries and regions than in others. For smaller countries, engagement is even more likely to be ad hoc.

A few lessons learned that EMBs and social media company representatives have shared with regard to establishing productive relationships:

- Both sides should establish clear communication channels and designated points of contact.
- Companies should establish relationships early in the electoral cycle when election authorities have capacity to engage and there is sufficient time to build trust.
- EMBs should take ownership by having an idea of what they want from social media companies and how they want to collaborate.
- EMBs should situate their coordination with social media companies within larger multi-stakeholder efforts as appropriate. For example, if an EMB is working with both social media companies and international implementers to optimize their use of social media, ensuring that these efforts reinforce one another can increase their value and reduce duplicate efforts.
- When desired, international implementers may facilitate or provide structure to the collaboration between an EMB and social media companies. In some instances, having a third party that understands how EMBs operate and what types of collaboration are more feasible for the social media company can increase the utility of these interactions and help EMBs feel confident that their interests are well represented.

Though there are similar services or types of coordination that social media companies provide across countries, the exact nature of coordination differs from country to country. As discussed, one fundamental distinction in EMB approaches to electoral disinformation is whether focus is on

enhanced dissemination of credible information or on sanctions for problematic content. This distinction informs the types of collaboration that an EMB is likely to engage in with social media and technology companies, though many EMBs will coordinate in ways that fall under both categories.

7.1 WORK TO HELP EMBs TO ENHANCE DISSEMINATION OF CREDIBLE INFORMATION



Google's Doodle for Indonesia Elections which ran April 17, 2019

EMBs may partner with social media and technology companies on a range of initiatives that expand the reach of EMB's public messaging or connect voters with credible electoral information.

Platform-embedded Voter Information

A common offering from Google and Facebook are Election Day reminders that direct users to EMB websites for additional details about how to participate in elections. In an increasing number of countries, Facebook will include Election Day notifications at the top of users' news feeds, which may include the ability to mark that you have voted in a way that is visible to your friends. In some countries, Google will alter the Google "doodle" (the changing image on the search engine's homepage) with an election-themed image that will link to country-specific voter information resources. In addition to Election Day notifications, [Facebook](https://www.facebook.com/help/1519550028302405) (<https://www.facebook.com/help/1519550028302405>) and [Google](https://elections.google/#protecting-elections) (<https://elections.google/#protecting-elections>) may also integrate notifications around voter registration deadlines, candidate information or details on how to vote. Google enabled an Informed Voting button one week before 2018 Mexican elections that redirected users to an INE microsite with information designed for first-time voters.² While platforms may run these notifications independently, in some countries companies will engage the EMB to verify that the information being provided is correct. Both Google and Facebook have also debuted tools to help voters find their polling locations – which either directs voters to EMB resources or relies on detailed data provided or verified by the EMB – such as a Google Maps integrated feature.

In addition to working with Facebook and Google, the TSE in Brazil pioneered a number of avenues for working with additional platforms. For example, the TSE partnered with WhatsApp to [develop a chatbot](https://www.techtudo.com.br/noticias/2020/11/eleicoes-2020-whatsapp-bane-mais-de-mil-contas-por-causa-de-fake-news.ghtml) (<https://www.techtudo.com.br/noticias/2020/11/eleicoes-2020-whatsapp-bane-mais-de-mil-contas-por-causa-de-fake-news.ghtml>) that answered election-related questions asked by users and helped them identify whether information was accurate. The chatbot also provided information on candidates and on when and where to vote. More than 1.4 million WhatsApp users queried that chatbot during the election period, and 350,000 accounts exchanged 8 million messages with the chatbot on Election Day alone. For the 2020 Brazilian elections,

Instagram created stickers to reinforce the importance of voting, automatically redirecting users to the TSE official website. Twitter created a notification for users with a link to the TSE webpage and promoted the dissemination of official TSE content on the platform. TikTok launched a page to centralize reliable information about the election.

It is important for companies to work with EMBs to ensure that they are prepared for the extra traffic to their site that may result from these notifications. A Facebook notification that urged Indonesians to check their voter registration status resulted in so much traffic to the election authority's website that it crashed.

Civic Engagement and Voter Education Support

In some countries, the platforms will engage in more complex civic engagement efforts that aim to extend the reach of credible and informative content. In Mexico, technology companies partnered with INE to expand the reach of civic and electoral information. Facebook amplified INE's call for citizens to choose the topic of the third presidential debate, and all three debates were streamed on the platform. INE also collaborated with Twitter using Periscope, Twitter's live video streaming application, to broadcast the three presidential debates, and encouraged national engagement around the debates with a series of customized hashtags. INE was also able to use a Tweet-to-Reply tool, which allowed users who retweeted INE messages on Election Day to opt-in to receive preliminary election results in real time.

Training on **how electoral authorities can optimize their use of Facebook for voter education and voter information** is another avenue for collaboration with EMBs. In Indonesia, Facebook provided these trainings to public relations departments in provincial and regional election offices. While Facebook provided guidance on topics such as how to make compelling videos, the value of identifying the right messenger for content, and other ways to use the platform for their goals, the company makes clear that they do not provide guidance on what content should be shared, merely how to share content effectively.

Depending on the mandate of the EMB and the specifics of collaborative agreements, social media and technology companies may also engage with election authorities to deploy **news literacy ad campaigns** or **trainings for electoral stakeholders on understanding and detecting disinformation**. Similar efforts might also be organized with other national stakeholders outside of the EMB, additional details about these types of interventions can be found in the guidebook section on [Platform Responses](https://counteringdisinformation.org/topics/platforms/0-overview-platforms). (<https://counteringdisinformation.org/topics/platforms/0-overview-platforms>)

7.2 WORK WITH EMBs TO RESTRICT OR SANCTION



HIGHLIGHT

PROBLEMATIC CONTENT

Social media companies also provide various avenues for election authorities to identify content that should be restricted or removed from social media platforms.

Account Verification and Security

An important, uncontroversial avenue of collaboration is providing election authorities with support for the expeditious removal of social media accounts that are falsely claiming to be or speak for the EMB. The existence of imitation accounts can be highly problematic, discrediting the electoral process and possibly sparking violence. For example, in the context of highly contentious 2018 Kenyan elections, a fake Twitter account declared Uhuru Kenyatta president prior to the official release of Presidential results, an incident that IFES field staff identified as a trigger for sporadic violence in opposition areas. Several fake accounts used the image of the Chairman of the Election Commission to announce incorrect electoral results or threaten violence against other members of the Election Commission.

EMB imitation accounts are common, and the identification and removal of these accounts is a service that major platforms are able to provide to EMBs of any size with relative ease, provided a trusted communication channel exists between the company and the EMB. A secretariat member of the EMB of Malawi reported that Facebook had been of assistance in taking down fake accounts ahead of elections. The Central Election Commission of Georgia has reported the same. In Georgia, several fake CEC Pages were discovered. Though the CEC judged that their impact was minimal, the Commission acted expeditiously to have the accounts taken down, both by contacting Facebook and by directly writing to the page administrators to desist, which was successful in several cases. The imitation pages had the potential to erode the credibility of the CEC, prompting decisive action.

CYBERHYGIENE AND INFORMATION INTEGRITY

An area of overlap where EMB's cybersecurity and cyber hygiene practices have implications for information integrity is the protection of official EMB social media accounts and other online channels of communication. When EMB communication channels are hacked and then used to disseminate false information, the impact is not only the immediate confusion that might cause, but also has the potential to undermine the EMB's ability to be a trusted communication channel in the future and undermine faith in the credibility and professionalism of the EMB more broadly.

“The fake CEC page discovered during the pre-election period, titled “Election Administration (CEC)” using the same profile and background pictures, would give unserious answers to people asking relevant questions...Our reputation and credibility were at stake as [this] is the goal of the disinformation itself.” – Interlocutor at the CEC of Georgia

Facebook, and possibly other platforms, express an active desire to have all EMB official Pages “blue check” verified on the platform. At a gathering of EMB commissioners and staff in South Africa in early 2020, they set up a booth that EMB representatives could visit throughout the conference to have their accounts verified, call attention to imitation accounts, and discuss other account security issues. Facebook reiterates basic account security protocols as part of account verification, including the enabling of two-factor authentication to make EMB social media accounts more secure.

Whitelisting to flag problematic content

Social media companies might also provide EMBs with an accelerated channel for reporting content that violates platform community standards. The major U.S.-based platforms maintain provisions that prohibit content that constitutes election interference, voter suppression and hate speech. In some instances, establishing a reporting channel is done through a more formal process. In others, it can happen on a more ad hoc basis.

Indonesia’s reporting process with Facebook, for example, was a formal arrangement, with a reporting process that was discussed and designed to fit Bawaslu’s needs. Facebook trained EMB staff on the platform’s community standards and content review process and provided Bawaslu with a dedicated channel through which they could report violations. Facebook and Bawaslu had a series of meetings to clarify Facebook’s content review policies in relation to local law and to establish a procedure for Bawaslu’s reporting process during the electoral period. This process included Bawaslu classifying content they identified as problematic, what local law the content breached, and the argument for why the content was in violation of that law. This was then submitted as an Excel spreadsheet on a weekly basis to Facebook. The [complaints referral and adjudication \(https://counteringdisinformation.org/topics/embs/6-disinformation-complaints-referral-and-adjudication-process\)](https://counteringdisinformation.org/topics/embs/6-disinformation-complaints-referral-and-adjudication-process) subcategory contains more details on this process. Although this formal process was carefully designed and adopted, a Bawaslu representative indicated that their reporting process with Facebook was not as expeditious as with other platforms. The Bawaslu representative indicated that their formal reporting process with YouTube resulted in a quicker removal of violating content.

In India (https://pib.gov.in/newsite/PrintRelease.aspx?relid=189494), the election commission convened social media platforms ahead of the election and, as part of the voluntary code of ethics, platforms “agreed to create a high-priority dedicated reporting mechanism for the ECI and appoint dedicated teams during the period of General Elections for taking expeditious action on any reported violations.”

Channels for flagging content to the platforms often form on a more ad hoc basis, particularly in countries with smaller populations in which the platforms lack a physical presence. If pre-existing relationships with the platforms do not exist, it may be too late to establish a clear process by the time elections are called. Merely establishing contact may be insufficient to lay a foundation for a productive exchange of information that benefits EMBs. A representative from the EMB of Mauritius reported that Facebook had sent representatives to meet with them ahead of 2019 elections and encouraged the EMB to report voter interference content for removal. However, when the EMB did identify content during the election that directed voters to the wrong polling locations and falsely alleged that ballots were being tampered with (clear violations of Facebook's community standards related to voter interference), the EMB was unable to reach anyone at Facebook to remove the content.

For EMBs that have at present only ad hoc communication with the platforms, greater systemization of the process for elevating concerns to the platforms would be valuable. The platforms should ensure that they have sufficient staff redundancies and reporting channels that a response is not contingent on the one or two individuals who initially established contact with the EMB.

Pre-certification of political advertisers

An unprecedented arrangement (<https://pib.gov.in/newsite/PrintRelease.aspx?relid=189494>) that the Election Commission of India made for 2019 elections was to require political advertisements to be pre-certified by the Media Certification and Monitoring Committee before they ran on social media. Candidates provided the details of their social media accounts to the election commission as part of the process of filing their nominations, and platforms were required to allow those accounts to only run advertisements that had been certified. In addition, certification was required for all election advertisements that featured names of political parties or candidates for the 2019 general elections. The platforms were also obligated to remove political advertisements that did not have certification upon notification by the ECI. It is hard to imagine the platforms complying with measures such as this in a country with a smaller market audience than India, or one in which the company was not physically present. This intervention is further discussed in the section on Legal and Regulatory Responses.

Enforcement of the Silence Period

The enforcement of a campaign silence or cooling period immediately prior to Election Day (as defined in local law) is another area where some EMBs have coordinated with social media platforms. Both Indonesia and India successfully gained compliance from social media companies that they would enforce the silence period. Other election authorities that express interest in having a similar arrangement have been less successful in gaining the platforms' compliance.

During the 48-hour silence period before Election Day, India's Voluntary Code of Ethics (<https://pib.gov.in/newsite/PrintRelease.aspx?relid=189494>) compels platforms to remove objectionable content within three hours of it being reported to them by the Election Commission.

The ban in Indonesia applied only to paid advertising, not to posts that were organically disseminated. Indonesia adopted an assertive enforcement approach to the silence period by issuing letters to each of the platforms outlining the provisions of the ban on campaign advertising during the blackout period. Letters indicated a willingness to use the existing criminal provisions in the law to enforce platform compliance. Facebook initially argued that the boundary between regular advertising and political advertising would be hard to discern. Bawaslu responded it was not their responsibility to resolve that tension and that it was incumbent on the platforms to ensure that they were in compliance with the law. Bawaslu speculated that the force of this edict led to a conservative interpretation of what constituted political advertising by the platforms, leading them to restrict a larger array of borderline advertising during the three-day silent period than they might otherwise have done. Bawaslu estimated, based on reports they received from the platforms, that the ban led to the rejection of approximately 2,000 ads across all of the platforms during the three-day silence period.

“Does a country have the boldness to threaten Facebook and YouTube to follow the guidelines? If they have that boldness, tech companies will consider the position.” – Commissioner Fritz Edward Siregar, The General Election Supervisory Agency of Indonesia (Bawaslu)



HIGHLIGHT

FEATURED INTERVENTION:

The MoU between Brazil's TSE and WhatsApp established a dedicated communication channel to allow the TSE to directly report WhatsApp accounts suspected of bulk messaging. The TSE then provided citizens with an online form to report illegal bulk messaging, and upon receiving those reports, WhatsApp would promptly launch an internal investigation to verify whether the reported accounts had violated WhatsApp terms and policies on bulk messaging and auto-messaging services. In which case, the accounts engaging in prohibited behaviors would be banned. During the 2020 electoral period,

the TSE received 5,022 reports of illegal bulk messaging related to elections, which led to the banning of 1,042 accounts.

The enforcement of a silence period is not something that the platforms have acted upon without being compelled by local authorities, and smaller countries are unlikely to have the clout to demand compliance. Other dimensions of campaign silence periods are discussed in the legal and regulatory section of this guidebook.

ELECTION MANAGEMENT BODY APPROACHES TO COUNTERING DISINFORMATION

8. EMB COORDINATION WITH CIVIL SOCIETY (/TOPICS/EMBS/8-EMB- COORDINATION-CIVIL-SOCIETY)

Election Management Bodies can coordinate with civil society to enhance the reach of their messaging or extend their capacity to engage in time and labor-intensive activities such as fact-checking or social listening. The ability to forge these types of partnerships will vary significantly based on the credibility, independence, and capacity of both EMBs and CSOs in a given country.

EMB-CSO collaboration can be formalized to varying degrees. For example, in advance of the 2019 Indonesian elections, Bawaslu signed a Memorandum of Action (MoA) with fact-checking CSO Mafindo and election oversight CSO Perludem, outlining the parameters of their planned coordination to counter disinformation and online incitement. In South Africa, the coordination between CSO Media Monitoring Africa and the IEC in the development of their disinformation complaints referral and adjudication process included a close working relationship but was not formalized. Though partnerships should be reviewed regularly to ensure they are still serving their intended goals, collaborative relationships can also be long-standing as opposed to being re-invented every electoral cycle; Perludem has had a cooperative agreement in place with the KPU since 2015 to aid with voter information efforts, among other things.

Collaboration between EMBs and CSOs requires a careful balancing act to maintain the credibility and perceived independence of both entities. For CSOs, a visible relationship with an EMB can legitimize and raise the profile of the work that they are doing, but it can also open them up to accusations of partiality or abdication of their role as watchdogs of government institutions.

In the case of Media Monitoring Africa, which played a critical role in the development and delivery of the IEC's disinformation complaints referral and adjudication process in South Africa, the involvement of the IEC in the effort gave the project visibility and credibility with donors and with the social media companies that were initially skeptical of the idea. This credibility in turn allowed MMA to raise sufficient funds to develop the project and provide their assistance to the IEC at no cost to the institution, which removed any financial relationship that could have called into question their impartiality. Prelude also has a policy to not receive money from EMBs, and the executive director, having formerly worked for Bawaslu, is careful to ensure that communication between her office and the election authorities is transparently conducted through formal channels.

At the same time, a visible relationship with an EMB can call into question the impartiality of a CSO. For example, Mafindo's fact-checking work includes addressing disinformation about Bawaslu and the KPU, which has opened them up to criticism for too heavily relying on official rebuttals from those institutions rather than independent verification of the claims being investigated. Prelude reports that the media will come to them for clarification on some election-related stories because they provide more expeditious responses than official sources, which has opened them up to accusations that they serve as a public relations department for the KPU.

8.1 COALITION BUILDING

EMB coordination with CSOs can simultaneously serve several goals including consensus building about disinformation as a threat to elections, coordination, and amplification of rebuttals and counter-narratives as well as transparency and accountability.

As discussed in the section on *Codes of Conduct and Codes of Ethics* (<https://staging.counteringdisinformation.org/topics/embs/3-emb-codes-conduct-or-declarations-principle-electoral-period>) and the section on *Disinformation Complaints Referral and Adjudication Processes* (<https://staging.counteringdisinformation.org/topics/embs/6-disinformation-complaints-referral-and-adjudication-process>), the act of consultation can create a foundation whereby an EMB begins to **build a network of actors that can work together to combat electoral disinformation**.

Bawaslu's engagement with CSOs, universities, religious organizations, and youth groups to establish their Declaration of Principles and consult on the definitions of prohibited content in electoral campaigns provided a foundation for Bawaslu's multi-stakeholder intervention strategies. The inclusion of religious leaders early, for example, meant a foundation for a relationship that could then help bolster the credibility of the EMB down the line, particularly in the context of the Hoax Crisis



HIGHLIGHT

In 2019, Brazil's TSE launched its "[Combating Disinformation Program](https://english.tse.jus.br/noticias-tse-en/2019/Setembro/tse-launches-the-)" (<https://english.tse.jus.br/noticias-tse-en/2019/Setembro/tse-launches-the->

Centers. Building broad coalitions of this nature is also something that INE in Mexico did ahead of the 2018 elections, bringing together civil society representatives, media, academics, political leaders as well as social media company representatives for a conference to discuss countering the influence of disinformation. This initial conference was then followed by coordination meetings (<https://www.ndi.org/our-stories/conference-mexico-tackles-issue-disinformation-ahead-elections-july>) between civic tech groups, fact-checkers, and citizen election observer groups to collaborate on their efforts to combat disinformation in the elections. In August 2019, Brazil's TSE launched its "Combatting Disinformation Program," which emphasized media literacy, after securing more than 40 institutional partners including media outlets, fact-checking agencies, and technology and social media company representatives.

program-against-disinformation-focused-on-the-2020-elections)" focused on November 2020 elections. The program brought together approximately 60 organizations including fact-checking organizations, political parties, education and research institutions and social media platforms.

The program organized efforts around six themes: TSE internal organization; training and capacity building; containment of disinformation; identification and fact-checking of disinformation; revision of the legal and regulatory framework; and improvement of technological resources.

The establishment of networks and coalitions can also help the EMB to **amplify voter information messages and messages to counter misinformation or incitement**. For example, part of the MoA outlining cooperation among Bawaslu, Marino, and Perludem included a joint information dissemination strategy to maximize each organization's network for better outreach. Besides, Perludem undertook voter information efforts in cooperation with the KPU to promote understanding of each phase of the voting process and the role of the EMB – a proactive communication tactic that can make it more difficult for voters to be deceived by misinformation and disinformation about the electoral process. They also worked with both election management bodies to integrate website features that allowed the networking of information among the EMBs, their own work, and the work of journalists. As part of this effort, they worked with the KPU to develop an API that they could use to directly pull official data from the KPU to populate the Prelude website. They also allowed disinformation reports from the public to be channeled to Bawaslu by integrating the Perludem website with CekFacta – a journalist fact-checking network.

Coordination with CSOs can also help promote the **accountability of Election Management Bodies**. For example, Prelude, in addition to providing a portal through which individuals could report disinformation complaints to Bawaslu also monitored the progress of the reports that were submitted through their system for an added level of transparency about how reports were being handled.

8.2 FACT-CHECKING AND COMPLAINTS REFERRAL

An EMB is unlikely to have the capacity or need to run its own fact-checking operation. However, having the EMB as an external contributor to a fact-checking operation can enhance the effectiveness of those efforts surrounding an election.

Establishing communication links with the EMB can enable fact-checking organizations to receive quick clarification in an instance where the EMB can authoritatively weigh in on the accuracy of a piece of false or misleading information in circulation.

INE had a role to play in the #Verificado2018 fact-checking effort in Mexico, which is discussed in greater detail in the chapter on civil society responses (<https://counteringdisinformation.org/node/2690>). The collaboration was particularly valuable on Election Day, as INE was able to quickly clarify several situations. For example, INE quickly filmed and shared a video explaining why special polling sites were running out of ballot papers in response to complaints coming from those polling sites. #Verificado2018 journalists also consulted INE to verify or rebut reports of election-related violence, with that information then widely disseminated via the media. INE's agreement with the #Verificado2018 team of journalists was that election authorities would provide clarification on every issue brought to them as soon as possible and that the Verificado team of journalists, in turn, would consult INE before publishing allegations, in addition to seeking confirmation through independent sources.¹

The arrangement between MAFINDO and Indonesian election authorities – both Bawaslu and the KPU – was also designed to facilitate quick clarification in instances where electoral misinformation or disinformation was brought to them by the fact-checking network. In practice, it was difficult at times to get speedy clarifications, an issue that MAFINDO attributed to inefficiencies in the internal flow of information that could result in receiving conflicting information from different individuals inside the EMBs.

Fact-checking organizations in Brazil were also dissatisfied with the speed and comprehensiveness of responses to requests for clarification that they directed to the TSE during the 2018 elections. The TSE reported (<https://www.poynter.org/fact-checking/2018/in-brazils-presidential-election-hoaxes-about-voter-fraud-run-rampant/>) that the volume of requests for clarification they received exceeded expectations and surpassed the capacity of their staff to respond.

From the perspective of programming support, coordination with external actors and clarifying internal lines of communication as part of strategic and crisis communication planning is something that could be of use. EMBs should be ready to be appealed to by fact-checking organizations, with a recognition that speed matters in responding. A communication protocol should clarify who should receive, process, and track the response to requests for information, who within the EMB has the authority to issue a clarification, and what the internal process for verifying the accuracy of information will be.

8.3 OUTSOURCING SOCIAL LISTENING

Like fact-checking, social listening to inform rapid incident response is another labor-intensive endeavor that EMBs may lack the capacity to conduct on their own. Civil society may be able to fill this gap through partnerships with EMBs.

In 2012 and 2016 independent media organization [Penplusbytes](http://penplusbytes.org/fighting-disinformation/) (<http://penplusbytes.org/fighting-disinformation/>) established [Social Media Tracking Centers](https://www.getaggie.org/) (<https://www.getaggie.org/>) (SMTC) to monitor social media during Ghanaian elections. The SMTCs used an [open-source software](http://africanelections.org/ghsmtc/about/) (<http://africanelections.org/ghsmtc/about/>) that presents trends in voting logistics, violence, political parties, and other topics. These were monitored for a continuous 72-hour period by Penplusbytes staff and university students. The process included a tracking team to monitor the social media environment and pass suspect content on to a verification team that would check the accuracy of the content forwarded to them. Problematic content was then sent to an escalation team that passed on information to the National Elections Security Task Force. Members of the SMTC were also embedded within the National Election Commission.

If setting up a dedicated effort like the SMTC's is not feasible, an EMB may be able to achieve many of the objectives of social listening for incident response through existing partnerships. Exchanging intelligence on trending narratives related to the election with fact-checking networks can be a means for EMBs to achieve the goals of social listening without the investment to build internal capacity to do this work. Similarly, EMB-established portals that allow the public to report problematic content for review, such as the [Real 411 initiative in South Africa](https://counteringdisinformation.org/interventions/real-411) (<https://counteringdisinformation.org/interventions/real-411>), can provide a crowdsourced approach to gain insight into problematic narratives circulating on social media.

ELECTION MANAGEMENT BODY APPROACHES TO COUNTERING DISINFORMATION

9. EMB COORDINATION WITH OTHER STATE ENTITIES (/TOPICS/EMBS/9- EMB-COORDINATION-OTHER-STATE- ENTITIES)

EMB COORDINATION
WITH OTHER STATE
ENTITIES



HIGHLIGHT

Elections are a flashpoint for misinformation and disinformation, but they are certainly not the only target of disinformation campaigns launched against democratic actors. Ensuring that state entities beyond the EMB have an active interest in monitoring, deterring, and sanctioning disinformation is crucial. Coordination with these other state entities during electoral periods can be essential for enhancing an EMB's ability to preserve electoral integrity in the face of misinformation and disinformation. Coordination among state entities can also align efforts and messaging to enhance efficiency and prevent the confusion of uncoordinated approaches. Coordination with other state entities can also be a valuable strategy for EMBs that have limited resources to dedicate toward counter-disinformation efforts.

A [detailed case study \(/interventions/real-411\)](#) of the State Electoral office of Estonia's establishment of an ad hoc **interagency task force for countering disinformation** in elections demonstrates the ways in which an election management body with limited staff and a restricted mandate can mount a comprehensive counter-disinformation response.

"The election management body in a small-scale system cannot rely on its own capability and has to gather other specialist institutions. This does not mean that the different nodes of expertise should act on their own but, rather, through the election management body as the main focal point." – Dr. Priit Vinkel, head of Estonia's State Electoral Office

9.1 ESTABLISHING AREAS OF RESPONSIBILITY AND LINES OF AUTHORITY

An EMB's counter-disinformation mandate must be considered in conjunction with the efforts of other state entities to promote information integrity. Ministries of Information, Digital Ministries, and Foreign Ministries, for example, might all have counter-disinformation mandates. State intelligence agencies, the police, courts, media and communication oversight bodies, anti-corruption bodies, human rights commissions, parliamentary oversight commissions, and others may also have a role to play.

Given how many entities may possibly be involved, in the electoral context, it is valuable to understand what different state entities are doing, and what effective collaboration might look like. It may be that the EMB wants to step into an authoritative role during the electoral period. This happened in Bawaslu's case; in advance of elections, it became clear that there was no institution in Indonesia with the authority to supervise hate speech and disinformation on social media during the electoral period.

“We asked ourselves a question – are we as Bawaslu brave enough to jump in, to supervise everything? ...We put ourselves in the hot seat.” Commissioner Fritz Edward Siregar, The General Election Supervisory Agency of Indonesia (Bawaslu)

Clarifying lines of authority can ensure that there is an authoritative voice in dealing with non-state entities, such as social media companies or political parties. Social media companies in particular are more likely to engage if the expectations and guidance they are receiving from state entities is aligned.

Coordination might take the form of a task force, a formal cooperative agreement, or a more ad hoc and flexible arrangement. The role of the EMB may be different depending on whether that arrangement is a standing body that takes special actions during elections, or whether it is a group convened specifically for the purpose of countering disinformation during elections. In the case of the former, an EMB may be seen as more of a resource partner to an existing body. In the case of the latter, the EMB may be leading the response.

In Denmark, efforts to organize a coordinated government response to online misinformation and disinformation included the establishment of an inter-ministerial task force (<https://um.dk/en/news/newsdisplaypage/?newsid=1df5adbb-d1df-402b-b9ac-57fd4485ffa4>), which had a special but not exclusive focus on elections. In Indonesia, Bawaslu, the KPU, and the Ministry of Communications and Information Technology signed a Memorandum of Action (<https://jakartaglobe.id/news/bawaslu-kpu-ministry-join-forces-fight-fake-news-ahead-regional-polls/>) before the 2018 elections and continued their cooperation during the 2019 elections. The agreement focused on coordinating efforts to supervise and manage internet content, coordinating information exchange among institutions, organizing educational campaigns, and promoting voter participation.

9.2 FACILITATING COMMUNICATION

Once institutions agree on a working arrangement, they must take steps to operationalize it. Focused discussions that delineate responsibilities and procedures for coordination can lay the groundwork for flexible and responsive communications, enabling rapid alignment and action when needed.

In Indonesia, Bawaslu held a series of face-to-face meetings with not only the Ministry of Communications and Information Technology but with the intelligence community, the army and the police to discuss guidelines, procedures and the relationship among their institutions. Part of those discussions included identifying which entity and which individuals within those entities had the authority to issue clarifications on which issues. After the formal relationship was established, the agencies communicated via a WhatsApp group that enabled quick responses and minimized formality that could hamper effective coordination.

To illustrate how communication worked, Bawaslu shared an example in which they encountered a social media post alleging that official army vehicles were being used as part of campaign activities. The Ministry of Communications and Information Technology had the tools to find the content and bring it to the group's attention but lacked authority to take action. The social media platforms moderation action was limited as they would be unable to determine whether the claim was true or not. The army had the information to prove that this claim was false but had no authority to flag the content for removal. By coordinating through their established WhatsApp group, all of the relevant parties were able to expeditiously identify and act on the issue – a feat that Bawaslu indicates would have taken more than a day if communication had been routed through formal communication channels.

The existing plan also enables institutions to speak with a joint voice in the case of serious allegations that might impact electoral integrity. In the case of the highly-publicized “seven containers hoax” which alleged that cargo ships full of pre-voted ballots had been sent to Jakarta, Bawaslu, the KPU, and the police held a joint press conference to clarify the situation. The process projected a united front in the effort to counter disinformation in the election.

9.3 MAINTAINING INDEPENDENCE

In coordinating with other state entities, maintaining the independence of the EMB will be of paramount importance. In countries where government ministries, intelligence agencies or other potential collaborators are aligned with a governing party or political faction, the EMB must make a judgement call on whether and how to collaborate with these institutions.

This decision may be made on a case-by-case basis. Though strong coordination existed among a number of entities in Indonesia, Bawaslu deliberately chose not to make use of the government's LAPOR! system, a platform that facilitates communication between the public and the government, including functionality for receiving reports and complaints from the public that could have been adapted for Bawaslu's disinformation complaints referral process. Though the platform was judged to be a technologically sophisticated tool that would have been of great use, after multiple discussions, Bawaslu ultimately decided not to use the channel given that using a tool associated with the ruling party might jeopardize their perceived independence.

“One of our considerations when we work with others is our impartiality” – Commissioner Fritz Edward Siregar, The General Election Supervisory Agency of Indonesia (Bawaslu)

9.4 INTEGRATE INTO PROACTIVE AND REACTIVE PROGRAMMING APPROACHES

Coordination with state agencies is something that can be or is naturally integrated into the proactive and reactive counter-disinformation strategies explored in other subcategories of this chapter.

Proactive Strategies

Proactive Communication and Voter Education Strategies to Mitigate Disinformation

Threats – coordination with other state agencies can be a useful way to amplify messages to larger audiences. For example, in instances where countries have credible public health agencies, partnering to communicate messages about how voting processes are changing as a result of COVID-19 can mitigate the risk that changes to election procedures could be subjects of disinformation.

Crisis Communication Planning for Disinformation Threats – Including other state agencies in crisis communication planning can build trust and working relationships that enable EMBs to get clarification and align messaging with other state entities in a crisis scenario.

EMB Codes of Conduct or Declarations of Principle for the Electoral Period – If codes of conduct are consultatively developed, the involvement of other state agencies may be beneficial to include from the outset. If codes of conduct are binding and enforceable, coordination as described under Disinformation Complaints Referral and Adjudication may be necessary.

Reactive Strategies

Social Media Monitoring for Legal and Regulatory Compliance – EMBs may or may not have authority to monitor social media for compliance or to enforce violations. In instances where the EMB shares this mandate with other institutions, clarifying the comparative mandates of each body and establishing how those entities will work together is essential.

Social Listening to Understand Disinformation Threats – A minority of EMBs will be positioned to establish their own social listening and incident response system. Ministries of Information, intelligence agencies, or campaign oversight bodies may, however, already have capacity to conduct social listening. It may be the case that an EMB is unable to preserve their independence and coordinate with these entities, but if it is possible, an EMB should consider establishing a channel through which information can be effectively relayed or staff from another state agency can embed with the EMB during sensitive electoral periods.

Disinformation Complaints Referral and Adjudication Process – For enforcement, an EMB will need to coordinate with relevant entities that may have jurisdiction over different complaints. This may include media oversight or regulatory agencies, human rights commissions, law enforcement, or the courts.

ELECTION MANAGEMENT BODY APPROACHES TO COUNTERING DISINFORMATION

10. PEER EXCHANGE AMONG EMBS ON COUNTER-DISINFORMATION STRATEGIES (/TOPICS/EMBS/10-PEER-EXCHANGE-AMONG-EMBS-COUNTER-DISINFORMATION-STRATEGIES)

With few precedents to emulate, dialogue and exchange among EMBS that are developing counter-disinformation approaches are particularly important. Exchange enables election authorities to learn from peers making similarly difficult decisions and adjustments.

[IFES' Regional Europe program has established a working group for EMBS \(https://counteringdisinformation.org/interventions/europe-and-eurasia-regional-election-management-body-working-group-social-media\)](https://counteringdisinformation.org/interventions/europe-and-eurasia-regional-election-management-body-working-group-social-media) dedicated to tackling the challenges presented by social media and disinformation in elections. The virtual launch of the working group in May 2020 gathered nearly 50 election officials from 13 countries in the Eastern Partnership and Western Balkans and provided a forum to discuss the challenge of electoral misinformation and disinformation during the COVID-19 pandemic. The working group provides EMBS with a platform for continued peer learning, skill-building, and developing best practices. The effort complements the launch of a global working group that brings together election authorities and social media companies planned by the Design 4 Democracy Coalition.

EMBS that have been leaders in developing counter-disinformation strategies are also passing on lessons to peer institutions in other countries. INE has shared exchanges with electoral authorities from Tunisia and Guatemala to learn from Mexico's counter-disinformation approach during elections. The Election Commission of South Africa hosted global experts and EMB representatives from across Africa in [March 2020 \(https://ewn.co.za/2020/03/04/iec-s-global-conference-focuses-on-potential-pitfalls-of-social-media\)](https://ewn.co.za/2020/03/04/iec-s-global-conference-focuses-on-potential-pitfalls-of-social-media) to share experiences mitigating the impact of social media on electoral integrity.

"We perceive the danger of disinformation, but a lack of information leaves us feeling like we don't have sufficient information, and the result is fear.... We need more information about the problem and to map credible sources of resources so that we don't have fear to use those resources."- Southern African EMB Representative

UNDERSTANDING THE GENDER DIMENSIONS OF DISINFORMATION

0. OVERVIEW - GENDER & DISINFORMATION (/TOPICS/GENDER/0-OVERVIEW-GENDER-DISINFORMATION)

Written by Victoria Scott, Senior Research Officer at the International Foundation for Electoral Systems Center for Applied Research and Learning

Around the world, women and people who challenge traditional gender roles by speaking out in male-dominated spaces—such as political leaders, celebrities, activists, election officials, journalists, or individuals otherwise in the public eye—are regularly subjected to biased media reporting, the spread of false or problematic content about them, and targeted character assaults, harassment, abuse, and threats. Any woman, girl, or person who does not conform to gender norms and who engages in public and digital spaces is at risk, although the public may be most familiar with this behavior directed toward women leaders. Women who hold or seek positions of public leadership often find themselves facing criticism that has little to do with their ability or experience—like the criticism typically encountered by men in those same positions—and instead face gendered commentary on their character, morality, appearance, and conformity (or lack thereof) to traditional gender roles and norms. Their representation in the public information space is often defined by sexist tropes, stereotypes, and sexualized content. While not a new challenge, this phenomenon is increasingly pervasive and has been fueled by technology. Although this type of online malice is often directed at women and lesbian, gay, bisexual, transgender and intersex (LGBTI) individuals in the public eye, any person who deviates from gender norms risks being exposed to this type of abuse.

For donors and implementers, understanding the intersection of gender and disinformation is imperative to designing and delivering comprehensive and effective programming to counter disinformation and hate speech and promote information integrity. Without considering the different ways in which women, girls, men, boys, and people with diverse sexual orientations and gender identities engage in the digital information environment and experience and interpret disinformation, donor and implementer efforts to counter disinformation will not reach the individuals who are among the most marginalized in their communities. The impact and sustainability of these interventions will therefore remain limited. **Analyzing disinformation through a gender lens is imperative to designing and implementing counter-disinformation programs in a way that both recognizes and challenges gender inequalities and power**

relations and transforms gender roles, norms, and stereotypes. This approach is necessary if donors, implementers, and researchers hope to effectively mitigate the threat of disinformation.

An increasing body of research and analysis explores the role of gender in disinformation campaigns, including the gendered impacts of disinformation on individuals, communities, and democracies. While this research presents a compelling case for funders and implementers to view information integrity and counter-disinformation programming through a gender lens, current programming is often limited to interventions to prevent or respond to online gender-based violence or to strengthen women's and girls' digital or media and information literacy. These are important approaches to strengthening the integrity of online spaces and responding to the information disorder (<https://www.coe.int/en/web/freedom-expression/information-disorder>), but a greater range of programming is both possible and necessary.

EXPLORE FURTHER:

This section of the guidebook is intended to be a resource to assist donors, implementers, and researchers to apply a gender lens when investigating and addressing information integrity and disinformation. It will also assist funders and practitioners in integrating gender throughout all aspects of counter-disinformation programming.

The section begins by briefly outlining why counter-disinformation programming must be viewed through a gender lens.

- EXPLORE GENDER CONSIDERATIONS IN COUNTER-DISINFORMATION PROGRAMMING



HIGHLIGHT

DISTINGUISHING ONLINE GENDER-BASED VIOLENCE AND GENDERED DISINFORMATION:

Gendered disinformation and online gender-based violence are concepts that are often conflated. According to the framing used throughout this guidebook, online gender-based violence can be considered a type of gendered disinformation (using gender to target the subjects of attack in false or problematic content), but gendered disinformation is broader than what online gender-based violence encompasses. Gendered disinformation reaches beyond gendered attacks carried out online to include harmful messaging that exploits gender inequalities, promotes heteronormativity,

and deepens social cleavages. One reason for the frequent conflation of these terms may be that discussions of gender and disinformation typically rely on examples of gendered disinformation that are also examples of online gender-based violence. For instance, a common example is fake sexualized content (like sexualized deepfakes and photoshopped images or edited videos placing a specific woman's face onto sexualized content). This example can be considered both online gender-based violence and gendered disinformation. However, there are also examples of gendered disinformation messages that are not necessarily categorized as online gender-based violence, for example sensationalized and hyper-partisan junk news stories designed to deepen existing ideological divisions and erode social cohesion.¹ These two phenomena intersect, and both threaten the integrity of the information environment and full and equal participation in political, civic, and public spheres. It is important for counter-disinformation programming to not only prevent and respond to these direct attacks of harassment and abuse considered under the label of online gender-based violence, but also to prevent and respond to influence operations that exploit gender inequalities and norms in their messaging.

¹There are differing definitions of the term "gendered disinformation," and a variety of perspectives on what constitutes gendered disinformation and whether or how it is distinct from online gender-based violence, abuse, or harassment. See e.g. review of existing definitions and distinctions in Jankowicz et al.'s [Malign Creativity: How](#)

[Gender, Sex, and Lies are Weaponized Against Women Online](https://www.wilsoncenter.org/publication/malign-creativity-how-gender-sex-and-lies-are-weaponized-against-women-online)
(<https://www.wilsoncenter.org/publication/malign-creativity-how-gender-sex-and-lies-are-weaponized-against-women-online>). As scholars and practitioners continue to develop their thinking in this emerging field, these definitions and perspectives continue to evolve.

(/TOPICS/GENDER/1-GENDER-CONSIDERATIONS-COUNTER-DISINFORMATION-PROGRAMMING#GENDERCONSIDERATIONS)

The section then defines the term “gendered disinformation” and the gender dimensions of disinformation in each of its component parts (actor, message, mode of dissemination, interpreter, and risk).

- EXPLORE THE GENDER DIMENSIONS OF DISINFORMATION (/TOPICS/GENDER/1-GENDER-CONSIDERATIONS-COUNTER-DISINFORMATION-PROGRAMMING#GENDERDIMENSIONS)

- Explore: Actors
- Explore: Messages
- Explore: Modes of Dissemination
- Explore: Interpreters
- Explore: Risks

The section closes with a look first at the current approaches to countering disinformation with gender dimensions and then at some promising new approaches for gender-sensitive counter-disinformation programming. While gender-sensitive programming and good practices are still emerging in the information integrity field, this section of the guidebook offers promising approaches based on known good practices in related fields. Specific examples of integrating gender into counter-disinformation interventions are also included throughout the guidebook’s thematic topics.

- EXPLORE CURRENT APPROACHES TO COUNTERING -GENDERED DISINFORMATION AND ADDRESSING GENDER DIMENSIONS OF DISINFORMATION
([HTTPS://COUNTERINGDISINFORMATION.ORG/CURRENT-APPROACHES-COUNTERING-GENDERED-DISINFORMATION-AND-ADDRESSING-GENDER](https://counteringdisinformation.org/current-approaches-countering-gendered-disinformation-and-addressing-gender))
- EXPLORE PROMISING APPROACHES TO GENDER-SENSITIVE COUNTER-DISINFORMATION PROGRAMMING
([HTTPS://COUNTERINGDISINFORMATION.ORG/PROMISING-APPROACHES-GENDER-SENSITIVE-COUNTER-DISINFORMATION-PROGRAMMING](https://counteringdisinformation.org/promising-approaches-gender-sensitive-counter-disinformation-programming))

UNDERSTANDING THE GENDER DIMENSIONS OF DISINFORMATION

1. GENDER CONSIDERATIONS IN COUNTER-DISINFORMATION PROGRAMMING (</TOPICS/GENDER/1-GENDER-CONSIDERATIONS-COUNTER-DISINFORMATION-PROGRAMMING>)

The onus of responding to and preventing gendered disinformation should not fall on the shoulders of subjects of gendered digital attacks, nor on those targeted or manipulated as consumers of false or problematic content.

Donors and implementers might wonder what makes gendered disinformation unique and different from other types of disinformation, why it is important to analyze the digital information landscape and any form of disinformation (regardless of whether it is specifically gendered disinformation) from a gender perspective, or why it is necessary to design and implement counter-disinformation programming with gender-specific considerations. Answers to these questions include:

- Disinformation that uses traditional gender stereotypes, norms, and roles in its content plays to entrenched power structures and works to uphold heteronormative political systems that maintain the political domain as that of cisgender, heterosexual men.
- The means of accessing and interacting with information on the internet and social media differs for women and girls compared with men and boys.
- The experience of disinformation and its impact on women, girls, and people with diverse sexual orientations and gender identities differs from that of cisgender, heterosexual men and boys.
- Disinformation campaigns may disproportionately affect women, girls, and people with diverse sexual orientations and gender identities, which is further compounded for people with multiple marginalized identities (such as race, religion, or disability).

In designing and funding counter-disinformation activities, donors and implementers should consider the variety of forms that gendered disinformation, and gendered impacts of disinformation more broadly, can take. Counter-disinformation efforts that holistically address gender as the subject of disinformation campaigns and address women and girls as consumers of disinformation provide for multidimensional interventions that are effective and sustainable.

1.1 WHAT ARE THE GENDER DIMENSIONS OF DISINFORMATION?

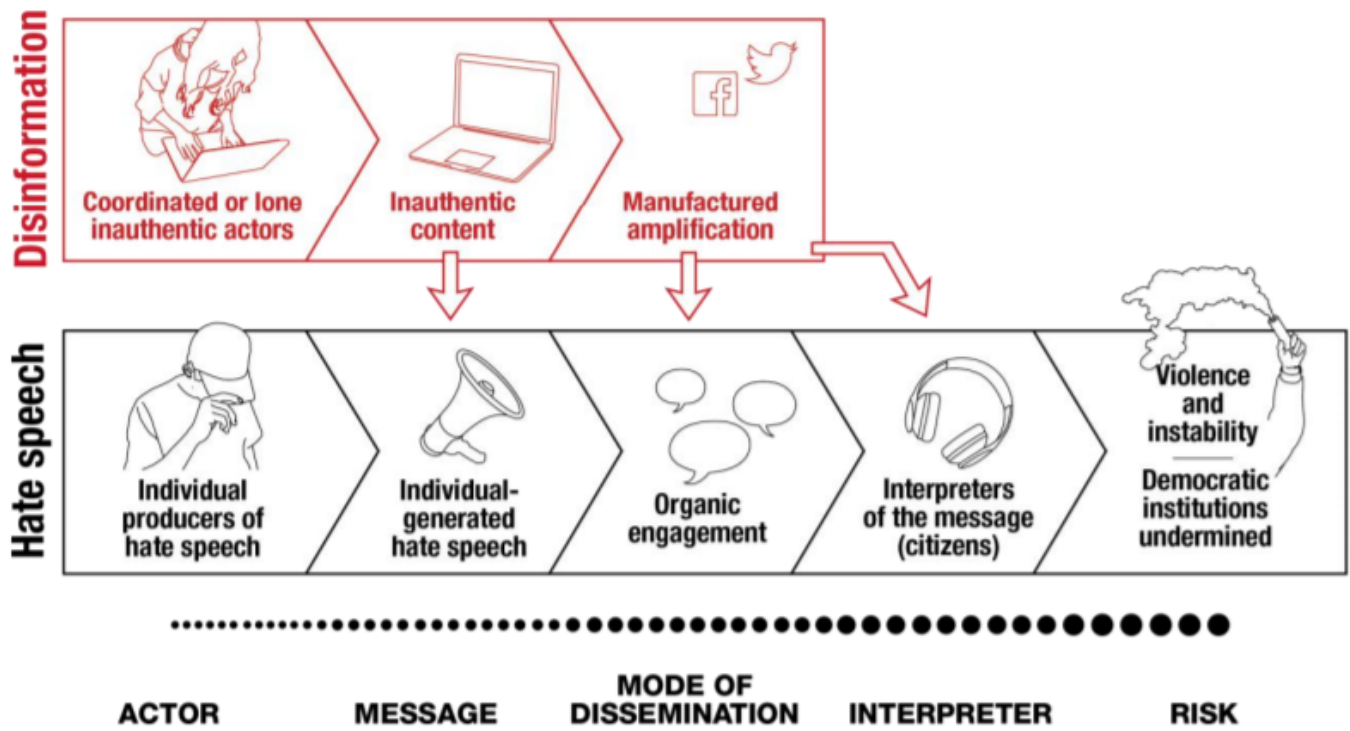
The intersection of information integrity challenges and gender is complex and nuanced. It includes not only the ways gender is employed in deliberate disinformation campaigns, but also encompasses the ways in which gendered misinformation and hate speech circulate within an information environment and are often amplified by malign actors to exploit existing social cleavages for personal or political gain. This intersection of gender and information integrity challenges will be referred to as “gendered disinformation” throughout this section.

Gendered disinformation includes false, misleading, or harmful content that exploits gender inequalities or invokes gender stereotypes and norms, including to target specific individuals or groups; this description refers to the content of the **message**. Beyond gendered content, however, other important dimensions of gendered disinformation include: who produces and spreads problematic content (**actor**); how and where problematic content is shared and amplified, and who has access to certain technologies and digital spaces (**mode of dissemination**); who is the audience that receives or consumes the problematic content (**interpreter**); and how the

creation, spread, and consumption of problematic content affects women, girls, men, boys, and people with diverse sexual orientations and gender identities, as well as the gendered impacts of this content on communities and societies (**risk**)¹.

By breaking down the gender dimensions of information integrity challenges into their component parts – actor, message, mode of dissemination, interpreters, and risk – we can better identify different intervention points where gender-sensitive programming can make an impact².

Below we illustrate the ways gender influences each of these five component parts of disinformation, hate speech, and viral misinformation.



Graphic: *The amplification of viral misinformation and hate speech through individual or coordinated disinformation*
 (https://www.ifes.org/sites/default/files/2019_ifes_disinformation_campaigns_and_hate_speech_brief.pdf)
 IFES (2019)

A. ACTOR

As with other forms of disinformation, producers and sharers of messages of disinformation with explicit gendered impacts may be motivated by ideology or a broader intent to undermine social cohesion, limit political participation, incite violence, or sow mistrust in information and democracy for political or financial gain. People who are susceptible to becoming perpetrators of gendered disinformation may be lone actors or coordinated actors, and they may be ideologues, members of extremist or fringe groups, or solely pursuing financial gain (such as individuals employed as trolls). Extrapolating from the field of gender-based violence, some of the risk factors (<https://www.cdc.gov/violenceprevention/intimatepartnerviolence/riskprotectivefactors.html>) that may contribute to a person's susceptibility to creating and spreading hate speech and disinformation that exploits gender could include:

- *At the individual level:* attitude and beliefs; education; income; employment; and social isolation
- *At the community level:* limited economic opportunities; low levels of education; and high rates of poverty or unemployment
- *At the societal level:* toxic masculinity or expectations of male dominance, aggression, and power; heteronormative societal values; impunity for violence against women; and patriarchal institutions

Gender-transformative interventions that seek to promote gender equity and healthy masculinities, strengthen social support and promote relationship-building, and increase education and skills development could build protective factors against individuals becoming perpetrators of gendered hate speech and disinformation. Similarly, interventions that seek to strengthen social and political cohesion, build economic and education opportunities in a community, and reform institutions, policies, and legal systems could contribute to these protective factors. In addition to identifying interventions to prevent individuals from becoming perpetrators of disinformation, practitioners must also acknowledge the complex discussions around the merits of [sanctioning actors for perpetrating disinformation and hate speech \(/topics/legal/0-overview-legal-and-regulatory-responses\)](/topics/legal/0-overview-legal-and-regulatory-responses).

It is worth noting that the present study did not identify any research or programming investigating women's potential role as perpetrators of disinformation. While it is widely known that the vast majority of perpetrators of online gender-based violence are men, researchers do not yet know enough about individuals who create and spread disinformation to understand whether, to what extent, or under what conditions women are prevalent actors. When considering the motivations and risk factors of actors who perpetrate disinformation, it is important to first understand who those actors are. This is an area that requires more research.

B. MESSAGE

Researchers and practitioners working at the intersection of gender and information integrity challenges have largely focused on the gender dimensions of disinformation messages. The creation, dissemination, and amplification of gendered content that is false, misleading, or harmful has been acknowledged and investigated more than other aspects of disinformation. The gendered content of disinformation campaigns typically includes messages that:

- Directly attack women, people with diverse sexual orientations and gender identities, and men who do not conform to traditional norms of "masculinity" (as individuals or as groups)
- Exploit gender roles and stereotypes, exacerbate gender norms and inequalities, promote heteronormativity, and generally increase social intolerance and deepen existing societal cleavages

There are myriad examples of disinformation in the form of direct attacks on women, people with diverse sexual orientations and gender identities, and men who do not conform to traditional norms of "masculinity" online. This can include sexist tropes, stereotypes, and sexualized content

(e.g. sexualized deepfakes or non-consensual distribution of intimate images³). Some of these cases—such as those targeting prominent political candidates and leaders, activists, or celebrities—are well-known, having garnered public attention and media coverage.

But while some cases of these attacks targeting prominent figures may be well-known to the public, many more cases of such gendered attacks online take place in a way that is both highly public and surprisingly commonplace. In 2015, [a report from the United Nations Broadband Commission for Digital Development's Working Group on Gender](#)



HIGHLIGHT

In 2016, leading up to the parliamentary elections in the Republic of Georgia, there was a disinformation campaign that targeted women politicians and a woman journalist in a video allegedly showing them engaged in sexual activity. The videos, which were shared online, included intimidating messages and threats that the targets of the attack should resign or additional videos allegedly featuring them would be released.

In another Georgian example, prominent journalist and activist, Tamara Chergoleishvili, was targeted in a fake video that allegedly showed her engaged in sexual activity with two other people. One of the people who appeared in the video with Chergoleishvili is a man who was labelled as “gay” and suffered consequences resulting from homophobic sentiments in Georgia.

Examples such as these seem sensationalized and extraordinary, but many women in the public eye encounter shocking instances of attacks like those described above. Similar cases of sexualized distortion have emerged against women in politics globally.

The potential impact of this type of gendered disinformation is to exclude and intimidate the targets, to discourage them from running for office, and to otherwise

disempower and silence them.

Perpetrators can also use these attacks to encourage their targets to withdraw from politics or to participate in ways that are directed by fear; to shift popular support away from politically-active women, undermining a significant leadership demographic, manipulating political outcomes, and weakening democracy; and to influence how voters view particular parties, policies, or entire political orders. Such attacks can also be used for gender policing (checking women and men who may be violating the gendered norms and stereotypes that govern their society).

Sources: Coda Story

(<https://www.codastory.com/disinformation/how-disinformation-became-a-new-threat-to-women/>), BBC

(<https://www.bbc.com/news/world-europe-35814185>), Radio Free Europe/Radio Liberty (<https://www.rferl.org/a/georgia-sex-tape-scandal-grigolia/27622049.html>)

(<https://en.unesco.org/sites/default/files/highlightdocumentenglish.pdf>) indicated that 73 percent of women had been exposed to or experienced some form of online violence, and that 18 percent of women in the European Union had experienced a form of serious internet violence at ages as young as 15 years. A 2017 Pew Research Center study (<https://www.pewresearch.org/internet/2017/07/11/online-harassment-2017/>) conducted with a nationally representative sample of adults in the U.S. found that 21 percent of young women (aged 18 to 29 years) reported they had been sexually harassed online. In a recently released 2020 State of the World's Girls report (<https://plan-international.org/file/46061/download?token=pH3r4scC>), Plan International reported on the findings from a survey conducted with more than 14,000 girls and young women aged 15-25 across 22 countries. The survey found that 58 percent of girls reported experiencing some form of online harassment on social media, with 47 percent of those respondents reporting that they were threatened with physical or sexual violence. The harassment they faced was attributed to simply being a girl or young woman who is online (and compounded by race, ethnicity, disability, or LGBTI identity), or backlash to their work and content they post if they are activists or outspoken individuals, "especially in relation to perceived feminist or gender equality issues." These direct attacks are not typically talked about

as unusual or surprising; rather, the risk of gendered attacks online is often considered a risk that women and girls should expect when choosing to engage in digital spaces, or—in the case of politically active women—“the cost” of doing politics (<https://www.ndi.org/not-the-cost>).

The contours of the digital information environment are characterized in part by this type of abuse, and these experiences have largely come to be expected by women and girls and tolerated by society. Though much of the time this content goes unreported, when survivors or targets of these attacks have brought complaints to law enforcement, technology companies and social media platforms, or other authorities, their concerns often go unresolved. They are commonly told that the content does not meet the standard for criminal prosecution (<https://www.mic.com/articles/114964/this-is-what-happens-when-you-report-online-harassment-to-the-police>) or the standard of abuse (<https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-4/>) covered by a platform’s code of conduct (</topics/platforms/0-overview-platforms>), advised to censor themselves, to go offline (or, in the case of minors, to take away their daughters’ devices), or told that the threats are harmless.

Beyond developing and deploying direct gender-based attacks against individuals or groups, disinformation actors may exploit gender as fodder for additional content. Such content may exploit gender roles and stereotypes, exacerbate gender norms and inequalities, enforce heteronormativity, and generally increase social intolerance and deepen existing societal cleavages. Examples include content that glorifies hypermasculine behavior in political leaders (<https://www.ndi.org/publications/engendering-hate-contours-state-aligned-gendered-disinformation-online>), feminizes male political opponents, paints women as being ill-equipped to lead or hold public office on the basis of gender stereotypes and norms, engages in lesbian-baiting, conflates feminist and LGBTI rights and activism with attacks on “traditional” families, and displays polarizing instances (real or fabricated) of feminist and LGBTI activism or of anti-women and anti-LGBTI actions to stoke backlash or fear. This type of content can be more nuanced than direct attacks and therefore more resistant to programming interventions.



HIGHLIGHT

Because of the ways that identity can be weaponized online, and the intersectional nature of gendered abuse, women, girls, and people with diverse sexual orientations and gender identities who also have other marginalized identities (such as race, religion, or disability) experience this abuse at higher rates and in different ways. (<https://medium.com/@AmnestyInsights/unsocial-media-tracking-twitter-abuse-against-women-mps-fc28aeca498a>)

C. MODE OF DISSEMINATION

Although gendered hate speech, viral misinformation, and disinformation are not new or exclusively digital challenges, the tools of technology and social media have enabled broader reach and impact of disinformation and emboldened those lone individuals and foreign or domestic actors who craft and disseminate these messages. Layering onto the range of harmful content that already exists in the information environment, disinformation campaigns designed to build upon existing social cleavages and biases can deploy a range of deceptive techniques to amplify gendered hate speech to make these gender biases seem more widely held and prevalent than they are.

Gendered hate speech and misinformation can have immense reach and impact even in the absence of a coordinated disinformation campaign, as this content circulates in the digital information space through organic engagement. While much of this content is generated and circulated in mainstream digital spaces, there is also a robust network of male-dominated virtual spaces, sometimes referred to collectively as the “[manosphere \(https://datasociety.net/wp-content/uploads/2017/05/DataAndSociety_MediaManipulationAndDisinformationOnline-1.pdf\)](https://datasociety.net/wp-content/uploads/2017/05/DataAndSociety_MediaManipulationAndDisinformationOnline-1.pdf),” where these harmful gendered messages can garner large bases of support before jumping to mainstream social media platforms. The “manosphere” includes online blogs and message and image boards hosting a variety of anonymous misogynistic, racist, anti-Semitic, and extremist content creators and audiences (“[men’s rights](https://www.splcenter.org/fighting-hate/extremist-files/ideology/male-supremacy),” “[involuntarily celibate](https://www.splcenter.org/fighting-hate/extremist-files/ideology/male-supremacy),” and other misogynist communities intersect with the “alt-right” movement in these spaces (<https://www.splcenter.org/fighting-hate/extremist-files/ideology/male-supremacy>))⁴.

Over time, the community of men who participate in these information spaces have developed effective strategies to keep these messages in circulation and to [facilitate their spread from anonymous digital forums with little moderation to mainstream \(social and traditional\) media \(https://arstechnica.com/gaming/2014/09/new-chat-logs-show-how-4chan-users-pushed-gamergate-into-the-national-spotlight/\)](https://arstechnica.com/gaming/2014/09/new-chat-logs-show-how-4chan-users-pushed-gamergate-into-the-national-spotlight/). Individuals who wish to disseminate these harmful messages have found ways to circumvent content moderation (such as using memes or other images, [which are more difficult for content moderation mechanisms to detect \(https://venturebeat.com/2020/12/01/ai-still-struggles-to-recognize-hateful-memes-but-its-slowly-improving/\)](https://venturebeat.com/2020/12/01/ai-still-struggles-to-recognize-hateful-memes-but-its-slowly-improving/)) and have developed tactics to inject this content into the broader information environment and to deploy coordinated attacks against specific targets



KEY RESOURCE

For definitions of some illustrative tactics, see [Defining \(https://onlineharassmentfieldmanual.pen.org/d/online-harassment-a-glossary-of-terms/\)](https://onlineharassmentfieldmanual.pen.org/d/online-harassment-a-glossary-of-terms/) “Online Abuse”: A Glossary of Terms - Online Harassment Field Manual (<https://onlineharassmentfieldmanual.pen.org/d/online-harassment-a-glossary-of-terms/>) by PEN America (<https://onlineharassmentfieldmanual.pen.org/d/online-harassment-a-glossary-of-terms/>) and [Online Abuse 101 \(https://www.womensmediacenter.com/speech-project/online-abuse-101/\)](https://www.womensmediacenter.com/speech-project/online-abuse-101/) by the Women’s

(individuals, organizations, or [mo](https://www.apc.org/en/pubs/facts-takebackthetech) (<https://www.apc.org/en/pubs/facts-takebackthetech>)).

[Media Center](https://www.womensmediacenter.com/speech-project/online-abuse-101/) (<https://www.womensmediacenter.com/speech-project/online-abuse-101/>).

This is in part what makes gender an attractive tool for disinformation actors. The “manosphere” provides ready-made audiences who are ripe for manipulation and activation in the service of a broader influence operation, and these communities have a toolbox of effective tactics for disseminating and amplifying harmful content at the ready. A known disinformation strategy includes the infiltration of existing affinity groups to gain group trust and seed group conversations with content intended to further a goal of the disinformation actor. Should disinformation actors manipulate these anti-women communities, they may successfully turn the energies of the “manosphere” against a political opponent, cultivating a troll farm with community members willing to carry out their work for free.

D. INTERPRETERS

Disinformation that targets women and people with diverse sexual orientations and gender identities as interpreters, or consumers or recipients, of disinformation is a tactic that can exacerbate existing societal cleavages – likely in ways that politically or financially benefit creators and disseminators of these messages. This can include targeting women and people with diverse sexual orientations and gender identities with disinformation designed to exclude them from public or political life (e.g., in South Africa, spreading false information that [people wearing fake nails or nail polish cannot vote in an election](https://citizen.co.za/news/south-africa/elections/2127918/people-who-have-fake-nails-can-vote-says-iec/) (<https://citizen.co.za/news/south-africa/elections/2127918/people-who-have-fake-nails-can-vote-says-iec/>)). In other cases, targeting these groups with disinformation may be part of a broader campaign to create polarizing debates and widen ideological gaps. For example, disinformation campaigns might inflame the views of feminists and supporters of women’s and LGBTI rights, as well as the views of those who are anti-feminist and who oppose women’s and LGBTI equality.



HIGHLIGHT

In November 2020, Facebook announced its takedown of a network of profiles, pages, and groups engaged in coordinated inauthentic behavior. The disinformation campaign, which originated in Iran and Afghanistan, targeted Afghans with a **focus on women as consumers** of the content shared. Almost half of the profiles on Facebook and more than half of the accounts on Instagram in the network were presented as women’s accounts. A number of pages in the network were billed as being for women. The women-oriented content shared across the network included a focus on content promoting women’s rights, as well as highlighting the Taliban’s treatment of women. The Stanford Internet Observatory’s analysis of the network indicated that additional content associated with the network was critical of the Taliban and noted that “[i]t is possible the intent [of

Disinformation that targets women and people with diverse sexual orientations and gender identities as interpreters of disinformation may amplify or distort divergent views to undermine social cohesion.

the women-focused content] was to undermine the peace negotiations between the Afghan government and the Taliban; the Taliban is known for restricting women's rights."

The potential impact of gendered disinformation like this is to deepen societal divides and exploit ideological differences, compromising social cohesion and undermining political processes.

Source: [Stanford Internet Observatory](https://cyber.fsi.stanford.edu/io/news/november-2020-takedowns) (<https://cyber.fsi.stanford.edu/io/news/november-2020-takedowns>).

E. RISK

The prevalence of technology and social media has brought new attention to the harms inflicted—especially on women—by information integrity challenges, including disinformation campaigns. Regardless of the individual motivations of the actors who create and disseminate gendered hate speech and disinformation, the gendered impacts of disinformation are typically the same:

- Exclusion of women and people with diverse sexual orientations and gender identities from politics, leadership, and other prominent roles in the public sphere through their disempowerment, discrimination, and silencing; and
- Reinforcement of harmful patriarchal and heteronormative institutional and cultural structures.

Harmful gendered content and messaging that seeks to deter women from entering political spaces and exploit social cleavages has become an expected, and in some cases accepted, part of the digital landscape. There are also implicitly gendered impacts of any form of disinformation campaign, as women may be the consumers or interpreters of any false and problematic content. Disinformation may also have a disproportionate effect on women and girls due to such factors as lower levels of



HIGHLIGHT

Gender-sensitive programming "attempt[s] to redress existing gender inequalities," while gender-transformative programming

educational attainment, media and information literacy, self-confidence, and social support networks, as well as fewer opportunities to participate in programming designed to build resilience against disinformation due to such factors as cultural norms and household and family care responsibilities. These are only a small sampling of the factors that likely cause women and girls to be disproportionately affected by disinformation, and result from broader gender inequalities such as unequal access to and control over resources, decision-making, leadership, and power. For this reason, effective counter-disinformation programming must address all aspects of the disinformation threat through designing and funding programming that is at minimum gender-sensitive, and ideally gender-transformative.

The gender dimensions of disinformation not only affect women and girls, but also people with diverse sexual orientations and gender identities, as well as people with other intersecting, marginalized identities. Due to limited relevant research and programming, there is minimal data available on this subject (a problem in and of itself), but members of the LGBTI population, as well as women and girls who have other marginalized identities, are targeted disproportionately by online harassment and abuse (<https://www.womensmediacenter.com/speech-project/research-statistics>) and likely also by disinformation campaigns (<https://www.thelily.com/black-women-are-being-targeted-in-misinformation-campaigns-a-report-shows-heres-what-to-know/>). It is imperative to consider the differential impact of disinformation on women, girls, and people with diverse sexual orientations and gender identities depending on other aspects of their identity (such as race, religion, or disability). They may be targeted in different ways in the digital information space than individuals who do not share these marginalized identities and may suffer more significant consequences from disinformation campaigns.

"attempt[s] to re-define women and men's gender roles and relations".

While gender-sensitive programming aims to "address gender norms, roles and access to resources in so far as needed to reach project goals," gender-transformative programming aims to "[transform] unequal gender relations to promote shared power, control of resources, decision-making, and support for women's empowerment.

Source: UN Women, [Glossary of Gender-related terms and Concepts](https://www.unwomen.org/en/digital-library/genderterm?AlphabetText=G) (<https://www.unwomen.org/en/digital-library/genderterm?AlphabetText=G>)

UNDERSTANDING THE GENDER DIMENSIONS OF DISINFORMATION

2. SIGNIFICANT GENDERED IMPACTS OF DISINFORMATION (/TOPICS/GENDER/2-SIGNIFICANT-GENDERED-IMPACTS-

DISINFORMATION)

The next two sections of the guide further explore two significant gendered impacts of disinformation:

- Silencing women public figures and deterring women from seeking public roles
- Undermining democracy and good governance, increasing political polarization, and expanding social cleavages

2.1 SILENCING WOMEN PUBLIC FIGURES AND DETERRING WOMEN FROM SEEKING PUBLIC ROLES

As the internet and social media have increasingly become major sources of information and news consumption for people across the globe, women in politics are turning to these mediums to reach the public and share their own ideas and policies as an alternative to often biased media coverage

(<https://static1.squarespace.com/static/5dba105f102367021c44b63f/t/5dc431aac6bd4e7913c45f7d>)

Many women—typically having limited access to funding, small networks, little name recognition, and less traditional political experience and ties than men in politics—note that their social media presence is integral to their careers and credit these platforms with giving them greater exposure to the public, as well as the ability to shape their narratives and engage directly with supporters and constituents. However, they also often find themselves the subjects of alarming amounts of gendered disinformation aimed at delegitimizing and discrediting them and discouraging their participation in politics.

According to research conducted by the Inter-Parliamentary Union

(<http://archive.ipu.org/pdf/publications/issuesbrief-e.pdf>) with 55 women parliamentarians across 39 countries, 41.8 percent of research participants reported that they had seen “extremely humiliating or sexually charged images of [themselves] spread through social media.” Not only do such experiences discourage individual women politicians from continuing in politics or running for reelection (either for concerns over their safety and reputation or those of their families), but they also have a deleterious effect on the participation of women in politics across entire societies, as women are deterred from entering the political field by the treatment of women before them.

“Research has shown that social media attacks do indeed have a chilling effect, particularly on first-time female political candidates. Women frequently cite the ‘threat of widespread, rapid, public attacks on their personal dignity as a factor deterring them from entering politics.’”

--(Anti)Social Media: The Benefits and Pitfalls of Digital for Female Politicians
(<https://static1.squarespace.com/static/595411f346c3c48fe75fd39c/t/5aa6fa310d9297a484994204/FINAL2-lowres.pdf>), Atlanta

Although there has been a recent increase in research investigating women politicians' experiences with gendered disinformation in the digital information space and social media⁵, this phenomenon is also experienced by women journalists, election officials, public figures, celebrities, activists, online gamers, and others. Women who are the subjects of disinformation, hate speech, and other forms of online attacks may be discriminated against, discredited, silenced, or pushed to engage in self-censorship.

What may be even more impactful is the pernicious effects of these disinformation campaigns on women and girls who witness these attacks on prominent women. Seeing how women public figures are attacked online, they are more likely to be discouraged and disempowered from entering the public sphere and from participating in political and civic life themselves. The subtext of these threats of harm, character assassinations, and other forms of discrediting and delegitimizing signals to women and girls that they do not belong in the public sphere, that politics, activism, and civic participation were not designed for them, and that they risk violence and harm upon entering these spaces.

2.2 UNDERMINING DEMOCRACY AND GOOD GOVERNANCE, INCREASING POLITICAL POLARIZATION, AND EXPANDING SOCIAL CLEAVAGES

"When women decide that the risk to themselves and their families is too great, their participation in politics suffers, as do the representative character of government and the democratic process as a whole."

--Sexism, Harassment and Violence against Women Parliamentarians
(<http://archive.ipu.org/pdf/publications/issuesbrief-e.pdf>), IPU

"Women's equal participation is a prerequisite for strong, participatory democracies and we now know that social media can be mobilized effectively to bring women closer to government – or push them out."

--Lucina Di Meo, *Gendered Disinformation, Fake News, and Women in Politics*

<https://www.cfr.org/blog/gendered-disinformation-fake-news-and-women-politics>

Beyond its impacts on women, girls, and people with diverse sexual orientations and gender identities as individuals and communities, disinformation campaigns that use patriarchal gender stereotypes or norms, use women as targets in its content, or target women as consumers undermine democracy and good governance. As scholar and political scientist Lucina Di Meo notes (<https://www.cfr.org/blog/gendered-disinformation-fake-news-and-women-politics>), inclusion and equal, meaningful participation are prerequisites for strong democracies. When disinformation campaigns hamper that equal participation, elections and democracies suffer.

Disinformation campaigns can use gender dimensions to increase political polarization and expand social cleavages simply by reinforcing existing gender stereotypes, magnifying divisive debates, amplifying fringe social and political ideologies and theories, and upholding existing power dynamics by discouraging the participation of women and people with diverse sexual orientations and gender identities. These actions serve to exclude members of marginalized communities from political processes and democratic institutions, and in so doing, chip away at their meaningful participation in their democracies and representation in their institutions. Because the voice and participation of citizens are essential to building sustainable democratic societies, silencing the voices of women, girls, and people with diverse sexual orientations and gender identities weakens democracies, making gendered disinformation not just a “women’s issue” and tackling it not just the mandate of “inclusion programming,” but imperative to counter-disinformation programming and efforts to strengthen democracy, human rights, and governance around the globe. A plurality of experiences and points of view must be reflected in the way societies are governed in order to ensure “participatory, representative, and inclusive political processes and government institutions.”

https://www.usaid.gov/sites/default/files/documents/1866/USAID-DRG_fina-6-24-31.pdf

UNDERSTANDING THE GENDER DIMENSIONS OF DISINFORMATION

3. CURRENT APPROACHES TO COUNTERING GENDERED DISINFORMATION AND ADDRESSING GENDER DIMENSIONS OF DISINFORMATION (/TOPICS/GENDER/3-

CURRENT-APPROACHES-COUNTERING-GENDERED-DISINFORMATION-AND-ADDRESSING-GENDER)

CURRENT APPROACHES TO COUNTERING GENDERED DISINFORMATION AND ADDRESSING GENDER DIMENSIONS OF DISINFORMATION

The field of gender-sensitive counter-disinformation programming is still emerging, and programming that explicitly centers the problem of gendered disinformation and gendered impacts of disinformation is rare. Currently, from the democracy to gender to technology sectors, there is limited, albeit growing, awareness and understanding of the nuanced and varied ways that disinformation and gender programming can intersect. To illustrate the variety of ways in which a gender lens can be brought to bear on counter-disinformation programming, programmatic examples that include gender elements are mainstreamed in the thematic sections of this guidebook. To complement these examples, this section applies what works in related programming areas to outline ways in which gender can be further integrated into counter-disinformation programming. For example, promising practices for gender-sensitive counter-disinformation programming can be drawn from good practices in development or humanitarian aid programs focused on gender-based violence and gender equity.

FOCUSED ON DIRECT ATTACKS OF ONLINE GENDER-BASED VIOLENCE

Existing programming to counter gendered disinformation is largely focused on preventing, identifying, and responding to direct attacks targeting women or people with diverse sexual orientations and gender identities as the subjects of gendered disinformation. These programs are often focused narrowly on women politicians and journalists as the targets of these attacks. This type of programming includes a variety of responses, such as [reporting and removal from platforms \(/topics/platforms/0-overview-platforms\)](#), [fact-checking or myth-busting \(/topics/csos/0-introduction-building-civil-society-capacity\)](#), digital safety and security training and skills-building, or media and information literacy for women, girls, and LGBTI communities. Similarly, the existing body of research identified as focusing on gendered disinformation is largely centered around diagnosing these direct attacks, the motivations of their perpetrators, and the harms of such attacks. While these are critical areas to continue funding for programming and research, these interventions are necessary but not sufficient. Donors and implementers must also pursue programming that addresses other dimensions of gender and disinformation.

To better inform the design and delivery of effective and sustainable interventions to counter gendered disinformation, as well as to mitigate the gendered impacts of disinformation more broadly, researchers must also broaden their focus to investigate such topics as:

- The different ways in which women, girls, men, boys, and people with diverse sexual orientations and gender identities engage with the digital information ecosystem
- The risk factors for and protective factors against perpetrating or being targeted by gendered disinformation
- Women as perpetrators of—or otherwise complicit parties to—disinformation, hate speech, and other forms of harmful online campaigns

Informative programming in this space might include digital landscape mapping, gender and technology assessments to identify gaps in access and skills, focus group discussions, community engagement, and public opinion research. This type of programming will enable practitioners to better understand the diverse ways in which these different groups interact with the digital information space, may be vulnerable to being targeted by disinformation or susceptible to perpetrating disinformation, and are affected by the impacts of disinformation.

MORE REACTIVE THAN PROACTIVE, MORE AD HOC THAN SYSTEMATIC

As noted in [other sections \(/topics/platforms/0-overview-platforms\)](/topics/platforms/0-overview-platforms) of the guidebook, one way to characterize counter-disinformation programming is to look at approaches as **proactive** or **reactive**.

Proactive programming refers to interventions which seek to prevent the creation and spread of gendered disinformation before it enters the digital information space. It might also include efforts to strengthen the resilience of those likely to be targeted by disinformation or those susceptible to becoming perpetrators of gendered disinformation. This can include a broad array of interventions, such as media and information literacy, confidence- and resilience-building, gender equality programming, civic and political participation programming, and education, workforce development, and livelihoods programming.

Reactive programming might include interventions which seek to respond to gendered disinformation after it has already been dispatched, such as reporting content to [platforms \(https://counteringdisinformation.org/topics/platforms/0-overview-platforms\)](https://counteringdisinformation.org/topics/platforms/0-overview-platforms) or [law enforcement for removal or investigation \(https://counteringdisinformation.org/topics/legal/0-overview-legal-and-regulatory-responses\)](https://counteringdisinformation.org/topics/legal/0-overview-legal-and-regulatory-responses) or fact-checking and responsive messaging to counter false or problematic content.

Some gender-sensitive counter-disinformation programming may be both reactive and proactive (https://counteringdisinformation.org/topics/platforms/0-overview-platforms), as they are interventions that both respond to the creation and spread of discrete cases of gendered

disinformation and aim to deter would-be perpetrators of gendered disinformation. Examples include platform- or industry-wide policies and approaches to identification, tagging, or removal of content, legislation to criminalize hate speech, online gender-based violence, and other harmful or problematic content, or regulation of platform responses to gendered disinformation.

Reactive approaches tend to be more **ad hoc** and immediate or short-term by nature, attempting to stamp out discrete disinformation campaigns or attacks as they emerge. Some proactive approaches are also ad hoc in nature, such as programs with one-off training sessions, classes, mobile games, or other toolkits for digital safety and security or media and information literacy. However, many proactive approaches (and some responses which are both reactive and proactive) are more **systematic** or long-term, aiming to transform gender norms, increase democratic participation, create long term social and behavior change, create safer spaces for women, girls, and people with diverse sexual orientations and gender identities online, and build the resilience of individuals, communities, and societies to withstand the weight of disinformation attacks and campaigns.

Much of the existing programming to counter gendered disinformation is reactive and ad hoc, designed to respond to gendered disinformation and address its impacts after it has already been pushed into the digital environment. Reactive interventions, such as content tagging or removal and fact-checking, myth-busting, or otherwise correcting the record in response to direct attacks, are generally insufficient to reverse the harms caused by gendered disinformation, from reputational damage and self-censorship to withdrawal from public and digital spaces and sowing seeds of distrust and discord.

As is the case with most gender-related programming, while there are important uses for both reactive and proactive programming to counter gendered disinformation, in order to ensure that disinformation prevention and response programming is both effective and sustainable, it is imperative that the donor and implementer communities think about proactive, not just reactive, gender-sensitive counter-disinformation programming. A major challenge, however, is that gender-transformative programming and programming designed to strengthen the protective factors against disinformation can typically be measured in generational shifts, rather than the two- to five-year periods most donor funding streams would require. Accommodating this holistic approach would require donors to consider rethinking the typical structure of their funding mechanisms and reporting requirements.



DESIGN TIP

However, as scholars and practitioners in this field will note, much of the damage has already been done by the time responses to gendered disinformation are deployed⁶.

UNDERSTANDING THE GENDER DIMENSIONS OF DISINFORMATION

4. PROMISING APPROACHES TO GENDER-SENSITIVE COUNTER-DISINFORMATION PROGRAMMING (/TOPICS/GENDER/4-PROMISING-APPROACHES-GENDER-SENSITIVE-COUNTER-DISINFORMATION-PROGRAMMING)

PROMISING APPROACHES TO GENDER-SENSITIVE COUNTER-DISINFORMATION PROGRAMMING

ESTABLISH INSTITUTIONAL AND ORGANIZATIONAL PROTOCOLS

Several recent research studies⁷ investigating the prevalence and impact of online harassment and abuse of (women) journalists in the United States and around the world have found that many subjects of such attacks do not report these incidents to their employers or other authorities out of concern that nothing can or would be done in response, or for fear of personal or professional repercussions from reporting. In cases where they do report these incidents to their employers, the organizations may not take action or may handle reports inconsistently and inadequately. A key recommendation that surfaced from these findings is to **establish institutional and organizational protocols, including specific policies and practices to support those attacked and to address reports of attacks.**

Based on this research and work in the area of online gender-based violence, donors and implementers should support institutions and organizations such as political parties or campaigns, EMBs, news and media outlets, and activist or advocacy organizations to establish comprehensive institutional protocols to prevent attacks and respond to reports, including:

- Providing appropriate digital safety and security training and education about online harassment
- Establishing clear and accessible reporting mechanisms that ensure the safety and protection of survivors of online violence and gendered disinformation, as well as their

ability to freely participate in digital spaces

- Ensuring systematic and consistent investigation of reports of attacks and referrals to appropriate authorities
- Establishing a variety of responses that institutions will offer to support their staff or members who are subjects of attacks (e.g. screening and documenting threats, reporting to platforms and/or authorities, coordinating counter-messaging, and sharing guidance and providing support to staff or members who choose to block or confront the perpetrators of their attacks)
- Providing appropriate resources and referrals following a report, such as physical security, psychological support, legal support, and personal information scrubbing services

In order to determine what protocols are needed, and to be responsive to the lived experiences of women and people with diverse sexual orientations and gender identities at work, programming should allow time and funding for institutions to survey their staff about their experiences and involve staff in decisions about the protocols, policies, and practices.

This approach can be adapted from the journalism and media industry to other organizations and institutions where gendered disinformation attacks are prevalent, installing policies and practices to ensure supportive, consistent, and effective responses to direct attacks. This intervention can contribute to combatting the impunity of perpetrators of gendered disinformation attacks, as well as the silencing, self-censorship, and discouragement to participate in the political or public spheres by the subjects of these attacks.

COORDINATE PREVENTION, RESPONSE, AND RISK MITIGATION STRATEGIES AND ESTABLISH APPROPRIATE CASE MANAGEMENT AND REFERRAL PATHWAYS

Gendered disinformation, much like gender-based violence, is a challenge which requires the involvement of stakeholders across multiple sectors and at multiple levels. Prevention and response efforts to address gendered disinformation depend on cooperation between the public and private sectors, including technology firms and media outlets (especially social media and digital communications platforms), law enforcement and justice authorities, civil society, psychosocial and mental health providers, and other health providers in cases where technology-fueled disinformation efforts may result in physical harm. Further, gendered disinformation risk mitigation efforts also depend on cooperation and information sharing between these stakeholders and international- and national-level policymakers (to inform legal and regulatory reform), civil society actors (to advocate for appropriate, effective, and sustainable interventions), the education sector (to inform curricula related to critical thinking and analytical skills, media and information literacy, digital safety), and the security sector in cases where incidents of gendered disinformation may be part of a coordinated campaign by malign foreign or domestic actors.

Donors and implementers should look to the robust experience of the humanitarian aid sector, specifically that of gender-based violence (GBV) prevention and response coordinators and service providers, to **develop a coordinated approach to gender-sensitive disinformation interventions**. Specifically, funders and implementers can adapt and draw guidance from the *Handbook for Coordinating Gender-based Violence Interventions in Emergencies* (https://www.un.org/sexualviolenceinconflict/wp-content/uploads/2019/06/report/handbook-for-coordinating-gender-based-violence-interventions-in-emergencies/Handbook_for_Coordinating_GBV_in_Emergencies_fin.01.pdf) and model national-level coordination networks and protocols on relevant elements of the approach detailed in this handbook to implement gender-sensitive responses to disinformation.

Two important elements of a coordinated approach to GBV interventions in emergencies to carry over when adapting this approach are **case management** and the establishment and use of appropriate **referral pathways**. Establishing appropriate case management in this scenario might entail: 1) the stakeholder who receives a complaint of gendered disinformation (for instance, a social media platform or local police) conducts a standard intake process with the person reporting; and 2) the stakeholder who receives the complaint or report uses an established referral pathway to refer the reporting party to a local civil society organization (for instance, local women's organizations that are experienced GBV service providers) for case management and additional referrals as appropriate. Referring the reporting party to an established case manager that is trained to work with targets or survivors of gendered disinformation and networked with the other stakeholders can streamline supportive services for the reporting party by establishing one primary point of contact responsible for interfacing with them. The case manager organization would be responsible for communicating the various response and recourse options available, providing referrals to appropriate service providers in the referral network and referring cases to appropriate members of the coordination network for follow-up, and (in cases of a direct attack) providing support to the target or survivor of the attack.

Establishing referral pathways in this scenario would involve identifying or establishing appropriate organizations or institutions responsible for different aspects of responding to reports of gendered disinformation, ensuring all coordination network organizations and institutions have access to the referral pathways, enabling them to receive initial reports of incidents and refer reporting parties to a local case manager organization, and case managers informing the reporting party about available services and avenues to pursue different interventions or recourse. If the reporting party gives permission, the case manager should also connect them with relevant services in the referral pathway.

Donors should consider supporting:

- A mapping or sectoral analysis of relevant stakeholders
- A convening of practitioners and experts to discuss the gendered disinformation landscape and needs
- Providing training and sensitization to law enforcement authorities, legal practitioners, and policymakers on gender, online and technology-facilitated gender-based violence, and disinformation

- The establishment of a coordination network that includes social media and digital communications platforms, law enforcement and justice authorities, civil society, psychosocial and mental health providers, and other health providers
- The development of clear roles and responsibilities of network members, for example establishing case manager organizations with support from civil society and governments
- The development of response protocols to guide the coordination, management, prevention, and response efforts of the network, including the development of a case management methodology and referral pathway

This intervention can contribute to the delivery of a holistic, survivor-centered approach to gender-sensitive counter-disinformation prevention and response programming, as well as combat impunity for perpetrators by institutionalizing a consistent and systematic approach of reporting claims to platforms and law enforcement authorities for investigation and recourse.

BUILD NETWORKS AND COMMUNITIES OF SUPPORTERS AND DEPLOY COUNTERSPEECH

“Don’t feed the trolls” is a common refrain of warning offered to those who find themselves the subjects of gendered disinformation. Experts used to think the best way to counter direct attacks targeting someone due to their gender and exploiting gendered norms and stereotypes was to simply ignore the attacks. Yet, recently, the dialogue around this issue has begun to evolve.

While some still advise not to “feed the trolls”—in other words, to simply ignore or to block, report, and then ignore the harmful content hurled at and about them online—others who work with the subjects of these attacks, as well as those who have themselves been the subjects of such attacks, have begun to acknowledge the shortcomings of this approach. They point to the empowerment that subjects of gendered disinformation and those who witness it may derive from speaking up and calling out the attacks (or seeing others do so), and the need for outing misogyny when it rears its head in digital spaces. Research conducted as part of the Name It. Change It. (<https://www.womensmediacenter.com/reports/name-it-change-it-the-womens-media-center-guide-to-gender-neutral-coverage-of-women-candidates-politicians-2012>) project also indicates that women politicians who directly respond to sexist attacks and call out the misogyny and harassment or abuse they face online (or when a third party does so on their behalf) are able to regain credibility with voters who they may have initially lost as a result of having been attacked (https://wmc.3cdn.net/b2d5a7532d50091943_n1m6b1avk.pdf).

It is important to clearly state that, while there are ongoing and evolving discussions on this topic about how best individuals can or ‘should’ respond to gendered disinformation, **it is not the responsibility of those who find themselves the subjects of such attacks to respond in any one way, if at all, nor to prevent the occurrence or take steps to mitigate the risks of these attacks. Those suffering gendered disinformation attacks should not be expected to shoulder the burden of solving this problem.** Rather, it is the responsibility of a variety of

stakeholders—including the technology platforms, government institutions and regulatory bodies, political parties, media organizations, and civil society—to establish and implement effective approaches and mechanisms to prevent and respond to gendered disinformation, as well as to work to address its root causes and to mitigate its long-lasting and far-reaching impacts.

Nevertheless, best practice adapted from gender-based violence response programming indicates that when the subject of gendered disinformation reports an incident, they should be presented with information on the available options for response and recourse and any potential benefits and further risks associated with those options.

One such possible response to gendered disinformation is counterspeech, which the Dangerous Speech Project defines as “any direct response to hateful or harmful speech which seeks to undermine it,” also noting, “There are two types of counterspeech: organized counter-messaging campaigns and spontaneous, organic responses

(<https://dangerousspeech.org/counterspeech/#:~:text=Counterspeech%20is%20any%20direct%20r>

Individuals who have been targeted by harmful content online might choose to engage in counterspeech themselves, or they might choose to enlist the support of their own personal and professional community or an online network of supporters to craft and deploy counterspeech publicly on their behalf or privately with messages of support (for example via email or on a closed platform). The effectiveness of counterspeech is difficult to measure, in part because those who engage in counterspeech may have different goals (ranging from changing the attitude of the perpetrator to limiting the reach of the harmful content to providing the subject of an attack with supportive messages). However, emerging research and anecdotal evidence indicates that crafting and deploying counterspeech (whether by the subjects of these attacks, their institutions or organizations, or a broader online community of supporters) is a promising practice in responding to gendered disinformation.⁸

A variety of positive outcomes to counterspeech have been referenced, including:

- delivering a sense of empowerment back to the targets of gendered disinformation attacks, allowing them to take back their narrative
- increasing the likelihood of positive, civil, or “pro-social” comments and/or decreasing the likelihood of negative, uncivil, or “anti-social” comments
- drowning out harmful content with supportive counterspeech, both on public social media posts and in private communications
- demonstrating to those sharing harmful content that their language or message is not accepted

Social media monitoring can play an important role in countering gendered disinformation, and can be linked to the coordination and deployment of counterspeech activities in response to gendered disinformation attacks.

Researchers, practitioners, and civil society actors are increasingly engaging in social media monitoring activities to inform their understanding of gendered disinformation, to identify entry points to disrupt gendered disinformation, viral misinformation, and hate speech, and to advocate

for laws or regulations that are responsive to the growing challenges of online gender-based violence and the spread of harmful gendered content online.

Social media monitoring in the context of gendered disinformation can be used to serve two primary functions:

- To listen to speech taking place across the digital information environment, monitor sentiment, and provide an important window into the creation, dissemination, and amplification of harmful content
- To monitor the adherence of political actors, media, and public institutions to legal and regulatory guidance and codes of conduct around disinformation and hate speech, and to monitor technology platforms' enforcement of their community standards, terms of use, or codes of conduct

An early step donors, researchers, and implementors should take is to create methodologies and tools to monitor social media and collect data on gendered disinformation, hate speech, and viral misinformation. These should be adapted to local contexts and applied in research and programming in order to mount an effective effort to counter gendered disinformation. In 2019, CEPPS released a social media analysis tool to monitor online violence against women in elections. The tool includes a step-by-step guidance on how to identify trends and patterns of online violence, including: identifying the potential targets to monitor (i.e. women politicians, candidates, activists); defining the hate speech lexicon to monitor; choosing which social media platforms to monitor; selecting the research questions; running the analysis using data mining software; and then analyzing the results. A full description of the step-by-step process can be found in [CEPPS' Violence Against Women in Elections Online: A Social Media Analysis Tool](https://www.ifes.org/sites/default/files/violence_against_women_in_elections_online_a_social_media_analysis_tool.pdf)

(https://www.ifes.org/sites/default/files/violence_against_women_in_elections_online_a_social_media_analysis_tool.pdf)

NDI has also developed a methodology for effectively scraping and analyzing such data in its reports "[Tweets that Chill \(https://www.ndi.org/tweets-that-chill\)](https://www.ndi.org/tweets-that-chill)" and "[Engendering Hate \(https://www.ndi.org/publications/engendering-hate-contours-state-aligned-gendered-disinformation-online\)](https://www.ndi.org/publications/engendering-hate-contours-state-aligned-gendered-disinformation-online)" with Demos through research in five countries. An essential step of the methodology is creating a lexicon in local languages of gender-based harassing language and the political language of the moment through workshops with local women's rights organizations and civic technology organizations.

Some of the key lessons from this research include:

- **Contextually- and linguistically-specific lexicons of online violence must be created and then evolve:** "Across all case study countries, workshop participants highlighted the fluid and evolving nature of language and brainstormed ways to account for this nuance in the study methodology. For example, NDI learned from the Colombia workshop that violent language in Spanish varied across Latin America, with both Colombia-specific and words from other parts of the region being used within the country. In Indonesia, religious words or phrases were used, complicating and heightening the online violence by invoking religious messages at the same time. In Kenya, workshop participants noted that a number of violent words/phrases that were in common usage in spoken Swahili, had not yet made it into

written text online on Twitter. These varied lessons point to the need for contextually- and linguistically-specific lexicons that can be continuously refreshed, modified, and implemented with human coders working alongside computer algorithms.” (excerpted from “Tweets that Chill (<https://www.ndi.org/tweets-that-chill>)”)

- **Attention to minority communities and intersecting identities is essential:** “Online [violence against women in politics] is varied and contextual, as it differs from country to country and culture to culture. However, it is also the case that the expressions used and impacts of online violence can vary significantly between and among communities within the same country. For this reason, it is important to intentionally include and consider historically marginalized communities among women (e.g. women with disabilities, LGBTI women, and female members of religious and ethnic minorities) when exploring the phenomenon of online [violence against women in politics]. During the Colombia workshop, female representatives from the deaf community shared that the violence they faced was not in text, but through the uploading of violent GIFs and/or video clips in sign-language. It was explained that this delivery mechanism was particularly effective in conveying threat and insecurity because, for the majority of the members of the deaf community in Colombia, sign language is their first language, and the targeting was therefore unmistakable. Understanding that the kinds of threats and modes of online violence can differ substantially when targeting different marginalized communities indicates that further work is required to create relevant lexicons.” (excerpted from “Tweets that Chill (<https://www.ndi.org/tweets-that-chill>)”)
- **Center Local Expertise:** “How gendered disinformation is framed and spreads across a network varies greatly according to context. Identifying or mitigating gendered disinformation cannot be successful without the central involvement and direction of local experts who understand the subtleties of how gendered disinformation may be expressed and where it is likely to arise and when. Platforms should support the work of local experts in identifying and combating gendered disinformation, for instance through the provision of data access or the trialing of potential responses through changes to platform design. Automated systems for identifying gendered disinformation are unlikely to have high levels of accuracy - though if employed, should be employed transparently and overseen by local experts.” (excerpted from “Engendering Hate (<https://www.ndi.org/publications/engendering-hate-contours-state-aligned-gendered-disinformation-online>)”)

The Legal and Regulatory chapter section 6.2 on building capacity to monitor violations (<https://counteringdisinformation.org/topics/legal/6-enforcement#CapacitytoMonitor>) and the Election Monitoring chapter (</topics/monitoring/0-overview-election-monitoring>) explore these

concepts further.

Seemingly in response to what many perceive to be a lack of adequate interventions by policymakers and technology platforms to address the problem of gendered disinformation, a variety of NGOs, civil society, and advocacy organizations have designed interventions to train likely targets of these digital attacks (as well as their employers and allies and bystanders) to develop and implement an effective counterspeech campaign, while others have established online communities of supporters who are ready to support the targets of these attacks with counterspeech efforts (among other supportive services such as monitoring the digital space where the attack is taking place and assisting the target of the attack in reporting the incident).

Counterspeech training examples:

- Tactical Tech's Gendersec Training Curricula (<https://en.gendersec.train.tacticaltech.org/>) on "Hacking Hate Speech" – a training workshop curriculum on how to set up an online support network, create textual and visual counterspeech content, and deploy a counterspeech campaign
- PEN America's *Online Harassment Field Manual* (<http://onlineharassmentfieldmanual.pen.org/>) – a training guide for journalists and writers on how to respond to online harassment and abuse, including building a community of supporters and developing counterspeech messages; includes guidance for employers on how to support staff experiencing online harassment, including through counterspeech

Online communities of supporters and counterspeech programming examples:

- Hollaback!'s HeartMob (<https://iheartmob.org/>) project – an online platform that has an at-the-ready network of supporters to respond to users' reports of online harassment and provide positive counterspeech (among other supportive services)
- TrollBusters (<http://www.troll-busters.com/>) – an at-the-ready network of supporters to respond to women journalists' reports of online harassment by providing positive counterspeech; includes monitoring the targets' social media accounts for continued attacks and to send continued counter-messaging (among other supportive services)

Funders and implementers should consider providing support to **scale up interventions like those referenced above for building communities of supporters and crafting and deploying effective counterspeech campaigns (/topics/surveys/0-executive-summary), including supporting the integration of these civil society interventions (/topics/csos/0-introduction-building-civil-society-capacity) into technology platforms.**

STRENGTHEN PROTECTIVE FACTORS AND
BUILD RESILIENCE OF INDIVIDUALS AND
COMMUNITIES

Because gendered disinformation is born of gender inequality and discriminatory norms, deterring its creation, dissemination, and amplification in the digital information environment will require donors and implementers to think beyond the perceived scope of counter-disinformation programming. As noted previously, programming to strengthen the protective factors and build the resilience of individuals, communities, and societies against gendered disinformation may not look like programming that donors and implementers typically think of as counter-disinformation interventions. This programming should not be limited to interventions to build the resilience of individual women, girls, and people with diverse sexual orientations and gender identities (although this is one important type of response), but should also include gender-transformative interventions which aim to **strengthen the resilience and protection of whole communities and societies against both perpetration and consumption of gendered disinformation.**

Programming to strengthen individuals', communities', and societies' protective factors against the threat of gendered disinformation (and disinformation more broadly), includes interventions spanning development sectors, such as programming to:

- promote gender equity and gender justice
- transform discriminatory and patriarchal gender norms
- strengthen social cohesion
- increase democratic participation and inclusion
- improve equitable access to quality education
- increase economic stability and improve economic opportunities
- build media and information literacy
- strengthen critical thinking, analytical, and research skills
- provide social support and confidence-building opportunities

Some who work at the intersection of technology, disinformation, and gender will caution that a focus on interventions such as media and information literacy, critical thinking skills, and confidence-building inappropriately places the responsibility of withstanding disinformation and its effects on individuals who are being adversely affected by it, rather than on the technology sector and policymakers to identify and institute effective solutions. The onus of responding to and preventing gendered disinformation should not fall on the shoulders of subjects of gendered digital attacks, nor on those targeted or manipulated as consumers of false or problematic content. However, in order to stamp out the problem of disinformation, gender-sensitive counter-disinformation efforts must include thinking holistically about building resilience to disinformation and designing programming to strengthen the resilience not only of individuals, but also of communities and whole societies. Regionally or nationally implemented media and information literacy curricula, for example, does not place the responsibility on individual students to learn to withstand gendered disinformation, but rather works toward inoculating entire communities against information integrity challenges.

Donors and implementers should work to integrate gender-sensitive counter-disinformation programming across development sectors, building these interventions into programming focused on longer-term social and behavior change to build the resilience of individuals,

communities, and societies to withstand the evolving problem of disinformation.

LEGAL AND REGULATORY RESPONSES TO DISINFORMATION

0. OVERVIEW - LEGAL AND REGULATORY RESPONSES (/TOPICS/LEGAL/0- OVERVIEW-LEGAL-AND-REGULATORY- RESPONSES)

Written by Lisa Reppell, Global Social Media and Disinformation Specialist at the International Foundation for Electoral Systems Center for Applied Research and Learning

The legal and regulatory frameworks governing elections vary significantly in how comprehensively they have adapted to the widespread use of the internet and social media in campaigning. While lawmakers in some countries have made strides to bring their legal and regulatory frameworks in step with an evolving information environment, other frameworks are largely silent on the topic of digital media. As the tactics of social media and technology-enabled information operations are increasingly adopted by political actors as standard campaign practices, the absence of legal and regulatory guidance that sets bounds on permissible campaigning behaviors becomes increasingly problematic.

Carefully crafted laws and regulations can inhibit political actors from using disinformation and other harmful or deceptive online practices for personal and political gain in ways that erode the health of the democratic information environment. At the same time, the adoption of overly broad legislation can have chilling implications for political and electoral rights. While legal and regulatory reform to adapt to the ways social media and technology have changed elections is essential, grounding that reform in comparative, global good practice can aid regulators in considering the challenges of regulating this area.

Though most countries have established norms and rules to govern the flow of information via print and broadcast media during campaigns and elections, the democratic principles that inform these laws and regulations – freedom of expression, transparency, equity, and the promotion of democratic information – have not been consistently extended to social media and online campaigning. Regulation, however, must do more than simply extend existing media oversight mechanisms to the digital world. Social media and the internet have altered the ways in which individuals encounter, interact with, and create political and electoral information, requiring lawmakers and regulators to adopt approaches consistent with this changed reality.

"[Our organization] look[s] at media content in the run up to elections – we look at print, broadcast, traditional media – all of which are clearly covered by electoral guidelines and processes. If we see something on radio or TV, there would be a means of recourse in our electoral commission to deal with that appropriately. What we saw with digital media... [it wasn't] covered by anything or anyone. It was a huge gap." — William Bird, Director of Media Monitoring Africa (South Africa)

National legislation governing the use of digital media during elections and campaigns has the potential to close loopholes currently being exploited by domestic actors to manipulate the information environment around elections. The use of disinformation for a political advantage during campaign periods constitutes more than the dissemination of false or misleading information. Disinformation campaigns are often directed by actors that leverage deceptive and coordinated behaviors online to distort public understanding, heighten social polarization and undermine trust in elections and democratic institutions. These campaigns are supercharged by the nature, scale, and networked capacity of new online systems and can have an outsized impact on the political participation, societal perception, and safety of women and other marginalized groups. To construct a network that deploys disinformation at scale often requires financial resources to not only develop and test messages but also finance the amplification of those messages. In the absence of specific political and campaign finance guidelines for the use of social media in campaigning, few limitations exist on what behaviors are permissible, even in instances where those behaviors would seem to constitute a clear violation of principles that exist elsewhere in the law.

Some countries are developing novel approaches for dealing with the use of social media in campaigns and elections, sometimes in ways that lack international precedent. The intent of this topical section is to detail, categorize, and discuss the implications of these emerging national-level legal, regulatory, and judicial decisions. This section draws from an analysis of the electoral legal frameworks of more than forty countries across six continents. Many of the laws and policies collected in this topical section have not yet been tested extensively in election contexts, so it may not always be clear which will succeed in advancing their intended goals.



HIGHLIGHT

Regulation that would change the behavior of foreign adversaries or significantly alter the global business practices of social media platforms is unrealistic goals for legal reform processes at the national level.¹ However, regulating the actions of domestic actors during electoral periods or discrete laws that create pressure on the ways platforms operate within a country are viable areas for reform. Such regulation also builds on the existing mandate of regulatory or judicial bodies to oversee the behavior of domestic actors during elections, including candidates and political parties.

EXPLORE: DEFINITIONS, COMPARATIVE EXAMPLES, AND ENFORCEMENT CONSIDERATIONS

This section of the guidebook is intended to be a resource for lawmakers contemplating the regulation of digital and social media in their own electoral legal frameworks, as well as for international donors and implementers that may be providing comparative examples in the process.

1. **DEFINITIONS (/topics/legal/1-definitions):** The content in this section begins with a discussion of key definitional considerations that lawmakers must address in the regulation of social media during elections and campaigns, as well as examples of how different countries have chosen to define these concepts. Depending on how these concepts are defined, they have the potential to significantly alter the scope and enforceability of law.

- Explore key **definitional challenges** for regulators:
 - What constitutes social or digital media? (/topics/legal/1-definitions#DigitalSocialMedia)
 - What is online campaigning? (organic vs. paid content) (/topics/legal/1-definitions#OrganicPaid)
 - Does the law distinguish among political, campaign, electoral, and issue advertising? (/topics/legal/1-definitions#PoliticalCampaignAds)
 - Who are the payers and paid entities in online campaigning? (/topics/legal/1-definitions#PayersandPaid)
 - What constitutes a digital or social media advertising expenditure? (/topics/legal/1-definitions#DigitalExpenditure)
 - Is there a timeframe during which expenditures must be disclosed? (/topics/legal/1-definitions#DisclosureTimeframe)
 - Why are definitions of fake news and disinformation problematic? (/topics/legal/1-definitions#FakeNewsProblematic)

2. **COMPARATIVE EXAMPLES:** The text then proceeds with comparative examples and analysis of measures taken in national-level law, regulation, and jurisprudence. It looks at measures to restrict online content and behaviors during campaigning and elections, as well as measures to promote transparency, equity, and democratic information. The examples that are included can be explored individually according to interest and need not be read consecutively. Examples are intended to provide comparative perspectives to inform legal and regulatory reform discussions, though the inclusion of an example does not constitute an endorsement of that approach.

- (A) Explore measures to **restrict online content and behaviors** (/topics/legal/2-measures-restrict-online-content-and-behaviors) *during campaigning and elections:*
 - (i) Measures directed at domestic actors:
 - Prohibit social media campaigning outside of a designated campaign period (/topics/legal/2-measures-restrict-online-content-and-

- behaviors#CampaignPeriod)
- Restrict online behaviors that constitute an abuse of state resources (/topics/legal/2-measures-restrict-online-content-and-behaviors#AbuseofStateResources)
- Set limits on the use of personal data by campaigns (/topics/legal/2-measures-restrict-online-content-and-behaviors#LimitPersonalData)
- Limit political advertising to entities that are registered for the election (/topics/legal/2-measures-restrict-online-content-and-behaviors#RegisteredEntities)
- Ban the distribution or creation of deepfakes for political purposes (/topics/legal/2-measures-restrict-online-content-and-behaviors#deepfakeban)
- Criminalize the dissemination of fake news or disinformation (/topics/legal/2-measures-restrict-online-content-and-behaviors#CriminalizeDissemination)
- (ii) Measures directed at social media and technology platforms:
 - Hold platforms liable for all content and require removal of content (/topics/legal/2-measures-restrict-online-content-and-behaviors#PlatformsLiable)
 - Prohibit platforms from hosting paid political advertising (/topics/legal/2-measures-restrict-online-content-and-behaviors#ProhibitPaidAdvertising)
 - Hold platforms responsible for enforcing restrictions on political advertisements run outside the designated campaign period (/topics/legal/2-measures-restrict-online-content-and-behaviors#PlatformsEnforceAdRestrictions)
 - Only allow platforms to run a pre-certified political advertisement (/topics/legal/2-measures-restrict-online-content-and-behaviors#PrecertifiedAds)
 - Obligate platforms to ban advertisements placed by state-linked media (/topics/legal/2-measures-restrict-online-content-and-behaviors#BanStateMedia)
 - Restrict how platforms can target advertisements or use personal data (/topics/legal/2-measures-restrict-online-content-and-behaviors#RestrictTargetAds)
- (B) Explore measures to **promote transparency** (/topics/legal/3-measures-promote-transparency-during-campaigning-and-elections) *during campaigning and elections*:
 - (i) Measures directed at domestic actors:
 - Require the declaration of social media advertising as a campaign expenditure (/topics/legal/3-measures-promote-transparency-during-campaigning-and-elections#DiscloseAdExpenditure)
 - Require registration of party and candidate social media accounts (/topics/legal/3-measures-promote-transparency-during-campaigning-and-elections#RegisterAccounts)

- Require disclosure and labeling of bots or automated accounts (/topics/legal/3-measures-promote-transparency-during-campaigning-and-elections#LabelBots)
 - Require disclosure of the use of political funds abroad (/topics/legal/3-measures-promote-transparency-during-campaigning-and-elections#DiscloseForeignFunds)
 - (ii) Measures directed at social media and technology platforms:
 - Require platforms to maintain ad transparency repositories (/topics/legal/3-measures-promote-transparency-during-campaigning-and-elections#MaintainAdRepository)
 - Require platforms to provide algorithmic transparency (/topics/legal/3-measures-promote-transparency-during-campaigning-and-elections#AlgorithmicTransparency)
- (C) Explore measures to **promote equity** (/topics/legal/4-measures-promote-equity-during-campaigns-and-elections) *during campaigning and elections*:
 - (i) Measures directed at domestic actors:
 - Cap party or candidate social media expenditures (/topics/legal/4-measures-promote-equity-during-campaigns-and-elections#CapExpenditures)
 - (ii) Measures directed at social media and technology platforms:
 - Require platforms to publish advertising rates and treat electoral contestants equally (/topics/legal/4-measures-promote-equity-during-campaigns-and-elections#EqualTreatment)
 - Compel platforms to provide free advertising space to candidates and parties (/topics/legal/4-measures-promote-equity-during-campaigns-and-elections#FreeAdvertising)
- (D) Explore measures to **promote democratic information** (/topics/legal/5-measures-promote-democratic-information-during-campaigning-and-elections) *during campaigning and elections*:
 - (i) Measures directed at domestic actors:
 - Require parties and candidates to issue corrections when party members or supporters share bad information (/topics/legal/5-measures-promote-democratic-information-during-campaigning-and-elections#CorrectBadInformation)
 - (ii) Measures directed at social media and technology platforms:
 - Require platforms to offer election authorities free advertising space for voter education (/topics/legal/5-measures-promote-democratic-information-during-campaigning-and-elections#FreeVoterEdAds)

3. **ENFORCEMENT (/topics/legal/6-enforcement)**: Thoughtful regulation means little if it is not accompanied by meaningful consideration of how that regulation will be enforced. A lack of realism about enforcement threatens to undercut the authority of the regulatory bodies

enacting reforms and may establish unrealistic expectations of what is achievable through regulation alone.

- Explore **considerations for enforcement**:
 - Establishing which state entity has an enforcement mandate (/topics/legal/6-enforcement#EnforcementMandate)
 - Building capacity to monitor for violations (/topics/legal/6-enforcement#CapacitytoMonitor)
 - Considerations for evidence and discovery (/topics/legal/6-enforcement#EvidenceandDiscovery)
 - Available sanctions and remedies (/topics/legal/6-enforcement#SanctionsandRemedies)

LEGAL AND REGULATORY RESPONSES TO DISINFORMATION

1. DEFINITIONS (/TOPICS/LEGAL/1-DEFINITIONS)

1.1 WHAT CONSTITUTES SOCIAL OR DIGITAL MEDIA?

The online media environment continues to evolve, and regulations that are crafted today to address specific elements of that environment may quickly become out of date. Regulators must consider the full range of internet-enabled communication tools to determine how broadly or narrowly to craft their guidance.

Writing in 2014, (<https://www.idea.int/es/publications/catalogue/social-media-practical-guide-electoral-management-bodies>) International IDEA

(<https://www.idea.int/es/publications/catalogue/social-media-practical-guide-electoral-management-bodies>) defined social media as “web or mobile-based platforms that allow for two-way interactions through user-generated content (UGC) and communication. Social media are therefore not media that originate only from one source or are broadcast from a static website. Rather, they are media on specific platforms designed to allow users to create (“generate”) content and to interact with the information and its source.”

In the intervening years, the social web has continued to evolve and definitions such as the above may no longer sufficiently capture the range of online activities that regulators wish to address. Analysis by the Knight First Amendment Institute at Columbia University of the [100 most popular social media platforms](https://knightcolumbia.org/content/top-100-the-most-popular-social-media-platforms) (<https://knightcolumbia.org/content/top-100-the-most-popular-social-media-platforms>) articulates the definitional complexities of

classifying social media. Capturing campaign activity that takes place on digital messaging applications, such as WhatsApp, Telegram, or Signal, or on subculture internet forums, for example, may require a broader definition than the one above. The role of search engines, online advertisement distributors, or ad-based streaming internet television in campaigning may also require greater definitional breadth.

Germany's 2020 Interstate Media Treaty (Medienstaatsvertrag – “MStV”) Law provides one of the more comprehensive definitions (<https://www.osborneclarke.com/insights/new-state-treaty-media-replace-treaty-broadcasting-create-legal-framework-changed-media-landscape/#:~:text=After%20the%20heads%20of%20government,after%20ratification%20by%20the>) of the range of activities it seeks to govern. The law introduces “comprehensive media-specific regulations... for those providers that act as gatekeepers for media content or services to disseminate it” such as “search engines, smart TVs, language assistants, app stores, [and] social media.” The law attempts to provide detailed definitions (<https://www.lexology.com/library/detail.aspx?g=e50f9bb5-95bb-4293-b3d5-3fc4d0b3e84c>) under the categories of media platforms, user interfaces, and media intermediaries.

Rather than referring to social or digital media, the electoral code of Canada refers to “online platforms,” defining them based on the salient feature being regulated by the code, namely that they sell advertising. Canadian law defines an online platform as “an Internet site or Internet application whose owner or operator, in the course of their commercial activities, sells, directly or indirectly, advertising space on the site or application to persons or groups.”²

Other jurisdictions will further restrict the social media or online platforms obligated to comply with a new law or regulation based on a specific criterion, such as the number of users. Germany's *Netzwerkdurchsetzungsgesetz* (NetzDG) law, for example, which requires companies to expeditiously remove illegal content from their platforms, applies only to internet platforms with at least 2 million users.³

In defining social media or digital media, drafters will want to consider:

- What array of online behavior does this law address? Does it include all websites that allow paid advertising or public comments, such as online news sites or blogs? Does it apply to digital messaging applications (i.e. WhatsApp)? Search engines? Internet advertising distributors?
- Is the intent of the law purely to regulate online *paid* activities taking place on social media? If so, should the definition be focused on online entities that run paid advertising?
- Are the obligations created in this law too burdensome for small social media companies in ways that will stifle competition due to the high costs of compliance? As such, should the law be limited to platforms that exceed a certain number of daily users or have a certain amount of revenue or market value?

1.2 WHAT IS ONLINE CAMPAIGNING? (ORGANIC VS. PAID CONTENT)

Legal and regulatory frameworks may wish to distinguish between “organic” and “paid” activities undertaken by the actor being regulated. Organic campaign content, for example, would be material shared by a party or candidate with their established social media audience who may or may not engage with or further disseminate that material. The reach of organic content is determined by the size of a candidate or campaign’s social media audience – i.e. those entities that have chosen to follow or engage with the social media actor in question – as well as the quality and appeal of the content that is being shared.

“Paid content” on the other hand is material for which the actor being regulated has paid to bring added visibility among audiences that may not have chosen to engage with that material. Different social media and digital platforms have different paid features to expand the reach of content, including but not limited to, the placement of advertisements or payment to prioritize content in users’ social media feeds or search engine results. If a party pays for the development of campaign messages or materials, even if they are then distributed through organic channels, that too may qualify as an expense that must be reported, as discussed in the following definitional section on “What constitutes a digital or social media advertising expenditure?”

This distinction is particularly pertinent in instances where there are restrictions on campaigning outside of a designated period. For example, a clear definition is needed to delineate what online behaviors are permissible before a campaign period begins or during an electoral silence period directly prior to the election.

Regulators in different countries have chosen to answer this question in different ways, with some determining that both paid and unpaid social media content constitutes online campaigning, while others determine that regulation pertains only to paid advertising.

Venezuela’s electoral legal framework, for example, stipulates that unpaid political expression on social media by candidates or parties is not considered campaigning.⁴ The Canadian framework acknowledges the complexity of enforcing campaign silence online by exempting “the transmission of a message that was transmitted to the public on what is commonly known as the Internet before the blackout period ... that was not changed during that period.”⁵ Similarly, 2010 guidelines from the National Electoral Commission of Poland prohibit any online activity that constitutes campaigning during the



HIGHLIGHT

In deciding where to delineate the boundaries of online campaigning, regulators might consider whether their primary intent is to regulate the activities of candidates and parties’ official pages or accounts or whether they wish to regulate the activity of any social media user

election silence period but allows content that was posted online before the start of the silence period to remain visible.⁶

In defining online campaigning, regulators will want to consider:

- Do they wish to distinguish between content that is disseminated through paid and unpaid means?
- Is only content shared by parties and candidates subject to regulation, or do stipulations pertain to a broader array of internet users that may post political content or purchase political or issue advertisements?
- What is the regulatory body's capacity to monitor and enforce campaign violations, and does this impact how narrowly or broadly online campaigning is defined?

1.3 HOW DOES THE LAW DEFINE POLITICAL ADVERTISING, CAMPAIGN ADVERTISING, AND ISSUE ADVERTISING?

Domestic law may take a broad or narrow approach to define the types of advertising that are subject to scrutiny. Clearly delineating the criteria by which online paid advertisements will be deemed to fall into a regulated category is essential for any regulation that, for example, attempts to place guardrails around permissible political advertising or requires specific disclosures related to online political advertising.

Electoral codes and social media platforms use varying definitions for "political advertising," "campaign advertising," "election advertising," and "issue advertising." These phrases do not have universal definitions, and establishing the definitional distinctions among these concepts is a familiar challenge from the regulation of offline campaigning as well. For both online and offline campaigning, subtle distinctions within these definitions can significantly alter the scope and impact of a law.

For countries that have designated campaign periods, "campaign advertising" and "campaign finance" are terms used to delineate activities and expenditures that occur during that designated period, while "political advertising" and "political finance" would include the activities and

engaging in campaigning. If the objective is to govern the official social media accounts of candidates and parties, monitoring all of the posts and activity of these accounts - paid and unpaid - is a more achievable goal given that only a discrete number of accounts will need to be monitored for compliance.

On the other hand, if regulation aims to impact all social media users posting political content, rather than just the official accounts of parties and campaigns, monitoring all organic posts from every social media user becomes impractical and at risk of selective or partisan enforcement. Focusing on paid advertising, particularly in countries where social media platforms' ad transparency reporting exists, makes oversight of all paid political advertising a more realistic goal.

expenditures of a party that take place outside of the campaign period or relate to the general operations of the party.

For the purposes of this section of the guidebook, “political advertising” will be used as an overarching term to refer to advertising that is placed by political parties, candidates, or third parties acting on their behalf, as well as any advertisements (regardless of who has placed the ad) that explicitly reference a political party, candidate, or election or that encourages a particular electoral choice. “Campaign advertising” will only be used when referencing measures that apply specifically to a designated campaign period.

The distinction is important, as some party expenditures – for example the placement of advertisements that serve a voter education purpose – might be considered political advertisements or campaign advertisements depending on the definitions used. If the definitions are indistinct, candidates and parties that conduct voter education outside of the campaign period may argue such messages are part of their normal course of business and not part of a campaign, opening a pathway for parties to circumvent campaign regulations.⁷

The phrase “issue advertising” is used in this section to capture a wider array of advertisements that reference social or political issues but do not explicitly reference a party, candidate, or election. Issue advertisements can be placed by any entity, whether they are expressly political or not. Countries that subject a broader array of online issue advertising to regulation may choose to do so in order to deter clandestine advertising with political, social, or financial goals, but which do not specifically name candidates or parties in an attempt to skirt regulation. A broad definition significantly expands the array of advertising that must then be subject to rules or review. [Facebook notes \(https://www.facebook.com/business/help/214754279118974?id=288762101909005\)](https://www.facebook.com/business/help/214754279118974?id=288762101909005) that for countries tracking issue advertisements, these can come from an array of advertisers including “activists, brands, non-profit groups, and political organizations.”

Attempts to regulate issue advertisements also raise freedom of expression considerations for civil society and advocacy groups. In Ireland for example, regulated activities include those “...to promote or oppose, directly or indirectly, the interests of a third party in connection with the conduct or management of any campaign conducted with a view to **promoting or procuring a particular outcome in relation to a policy or policies** or functions of the Government or any public authority.”⁸ The debate over this provision highlighted concerns that such a broad definition could impact the advocacy and campaigning work of civil society organizations.⁹

New Zealand and Canada have also crafted sufficiently broad definitions of election advertising to make possible the sanction of online political advertising disguised as issue-based advertising.

- [New Zealand](#)¹⁰
 - *In this Act, **election advertisement**—*
 - *(a) means an advertisement in any medium that may reasonably be regarded as encouraging or persuading voters to do either or both of the following:*
 - *(i) to vote, or not to vote, for a type of candidate described or indicated by reference to views or positions that are, or are not, held or taken (whether or not the name of*

the candidate is stated):

- *(ii) to vote, or not to vote, for a type of party described or indicated by reference to views or positions that are, or are not, held or taken (whether or not the name of the party is stated);*
- Canada¹¹
 - **Election advertising** means the transmission to the public by any means during an election period of an advertising message that promotes or opposes a registered party or the election of a candidate, including by taking a position on an issue with which a registered party or candidate is associated.

Both New Zealand and Canada's definitions further distinguish election advertising from an editorial or opinion content.

Whether national law provides that advertisements about political or social issues are subject to additional transparency or oversight measures may impact the information that is collected and cataloged by Facebook, and possibly by other online platforms. For example, as of early 2021, Facebook captured a greater range of advertisements in its Ad Library for Canada, the European Union, Singapore, Taiwan, the United Kingdom, and the United States than for other countries.¹² Among the 34 countries that gained access to the Facebook Ad Library in July and August of 2020, only New Zealand and Myanmar required added disclosure for social issue advertising in addition to political and electoral ads (which applied for all of the remaining countries). In New Zealand's case, this may have been in response to a national-level legal provision requiring broader disclosures from the platform related to issue advertising, though Myanmar's legal code is silent on the topic.

In defining political, campaign, election or issue advertising, regulators will want to consider:

- Are there definitions of political, campaign, electoral, or issue advertising in the current electoral legal framework? If so, do they apply to social media advertising?
- If there is no definition, or it does not apply to social media, or it includes a narrow definition of political advertising, would it be beneficial to expand or revise the definition?
- Does the legal framework require activists, brands, non-profit groups and political organizations to disclose issue advertisements?
- In each instance, is this a reasonable burden to place on these entities that will not suppress their ability to reach intended audiences due to overly-onerous requirements?
- What are the implications of any proposed changes on freedom of expression, particularly for civil society organizations engaged in advocacy?

1.4 WHO ARE THE PAYERS AND PAID ENTITIES IN ONLINE CAMPAIGNING?

If regulators are attempting to use existing legal mechanisms at their disposal – including the legal framework regulating political finance, public corruption, or the use of state resources – then definitions that acknowledge the complexity of the information ecosystem need to be considered.

The creation of disinformation at scale by a domestic or foreign actor will likely necessitate the outlay of funds to secure the personnel, expertise, and materials needed to create and maintain a sustained online campaign. Regulation that seeks to bring transparency through disclosure requirements or regulate paid campaign activities must therefore acknowledge the multitude of financial relationships that might constitute an expenditure.

Digital and social media campaigning increases the opportunities to obscure the origins of content by acting through third parties. Measures that seek to bring transparency into these financial flows will want to consider not only who is the payer and beneficiary, but also who is the paid entity -- the social media platform itself?

Influencers who operate pages or feeds on respective platforms and may be paid to promote political content? Public employees, who engage in campaigning via social media while at work? Public relations firms or content creation entities (such as content farms or troll farms) that produce and disseminate content on behalf of a political entity?

Additionally, are those entities operating from within the country or extraterritorially?

Social media has changed campaigning, but some behaviors that a reasonable observer might deem problematic do not, in fact, constitute a campaign violation. This can result from a lack of definitional specificity about which paid activities constitute campaigning.

In Guinea, for example, researchers at the Stanford Internet Observatory turned up evidence of a network of Facebook Pages run by Guinean president Alpha Condé's political party. The network, using false accounts, deployed coordinated efforts to amplify party propaganda while obscuring their paid links to the party. The Stanford researchers observed that "[t]he Guinea case raises broader questions about where and how to draw the line between modern political campaigning in the age of social media and coordinated inauthentic behavior."

Canada, for instance, exempts social media posts from its definition of "advertising" if it falls within the following parameters: "the transmission by an individual, *on a non-commercial basis* on the Internet, of his or her personal political views" (emphasis added).¹³ This can be interpreted to require the payment of social media intermediaries or influencers by political entities to be disclosed as advertising. Without this consideration, candidates and political parties can circumvent regulations by paying third-party entities to promote content or place advertisements on their behalf. The nature of social media makes it comparatively easy for a political entity to engage the services of a third party to perform otherwise regulated or prohibited activities on social media while circumventing disclosure requirements. Laws should include clear definitions of terms to capture this reality and close loopholes.

Conversely, measures that sanction or place obligations on the disseminators of unlawful content – without seeking to identify the funders of that content – are unlikely to deter the actors that are the ultimate beneficiaries of disinformation campaigns.

In defining who the payer and paid entities are, regulators will want to consider:

- If a certain action is prohibited or subject to disclosure requirements, does the legal and regulatory framework also apply to the hiring or instruction of third parties to perform that action?
- How does the legal or regulatory provision under consideration impact the disseminator of content versus the funder of the activity?

There is no regulation to catch the funder, only the one who spread[s the content].”
— Indonesian Civil Society Representative

1.5 WHAT CONSTITUTES A DIGITAL OR SOCIAL MEDIA ADVERTISING EXPENDITURE?

If a legal or regulatory approach includes disclosure or transparency requirements, it is important to define the types of expenditures on digital advertisements or digital campaigns that must be disclosed. These requirements may also need to be reviewed at regular intervals to ensure that they suit the rapidly evolving tactics of digital campaigning.

Robust disclosure requirements will provide insights into the sources of funding, the amount of funding provided by each source, and detailed information on how funding was used. Full disclosure is necessary to make it possible to judge if funds are coming from legally allowable sources and being used for legitimate party and campaign purposes. Minimal disclosure requirements make it easy for political actors to comply with the letter of the law while concealing questionable behaviors that violate the intent of disclosure requirements.

Analysis by the UK Electoral Commission notes that digital advertising expenditures can be easily hidden under different reporting categories. The Commission notes that they are unable to capture an accurate picture of how much has been spent on social media advertising because data is limited to payments made directly by the reporting entity to identifiable social media providers, such as Facebook or YouTube. This does not account for the reality that a significant amount of digital spending happens via consultancies or intermediary advertising agencies.¹⁴ For example, the Labour Party reported digital advertising expenditures of £16,000 in the 2015 Parliamentary Elections in the UK, when later calculations showed the total to be closer to £130,000 via intermediary advertising agencies. Practices such as this led the Electoral Commission to conclude that more detailed expenditure requirements were needed.¹⁵

In defining what information to include in disclosure requirements, regulators will want to consider:

- What constitutes an expenditure? For example:
 - Only the cost to place an ad?
 - The payment of digital advertising or public relations firms to design and deploy ad campaigns?
 - The cost to produce an ad?
 - The cost to profile target audiences?
 - The cost to develop and deploy chatbots (or other bots) to engage with users on social media platforms?

- The direct or third-party payment of content (or troll) farms to disseminate designated social media content or messages in large numbers?
- The cost of obtaining influencer endorsements?

1.6 IS THERE A TIMEFRAME DURING WHICH EXPENDITURES MUST BE DISCLOSED?

For countries that have defined campaign periods outlined in law or regulation, a loophole opens if regulators require detailed disclosure of social media advertising expenditure only during the campaign period. Though such spending may still be captured in regular party financial reporting, figures might only be captured annually and, depending on reporting requirements, may contain less detail than what may be required during campaign periods. Additionally, whether an expense is defined as an agreement to make a payment or a payment itself can impact reporting. If imprecisely defined, a political contestant could, for example, delay payment to a social media intermediary until after Election Day to skirt reporting requirements.

In defining a timeframe for disclosure, regulators will want to consider:

- How are disclosure requirements already outlined in the law for traditional media or political finance?
- Is the timeline crafted in a way that aligns with when digital or social media expenditures are likely to take place in the electoral cycle? For example, the cost to profile target audiences or pay for an influencer endorsement could occur well in advance of the electoral event, or payment could happen after Election Day as a way to avoid disclosure requirements that cover only the immediate campaign period.

Campaign finance limits to expenditure only apply during the campaign period – but there are campaign expenditures also outside of the official campaign period... We have to redefine campaign finance coverage to be more comprehensive. — Indonesian Civil Society Representative

1.7 WHY ARE DEFINITIONS OF FAKE NEWS AND DISINFORMATION PROBLEMATIC?

Legal and regulatory interventions that attempt to ban or sanction “fake news” or disinformation have been widespread in recent years. However, the difficulty defining these terms is one of the reasons such measures are frequently criticized by those who fear their implications for fundamental rights. As discussed in [the introduction to this guidebook](https://counteringdisinformation.org/introduction) (<https://counteringdisinformation.org/introduction>), precise definitions of disinformation are elusive, and what is commonly referred to as disinformation encompasses a wide range of deceptive and problematic behaviors.

If the success of a legal or regulatory intervention relies on a precise, comprehensive, and universally applicable definition of “fake news,” “false information,” “misinformation,” “disinformation,” or a similar term, it is likely that the intervention will either result in collateral damage to freedom of expression or be too vague to be reliably enforceable. It also holds a high risk of being selectively enforced, for example, against political opponents or to restrict press freedoms.

Some jurisdictions have chosen to leave the issue of determining what content constitutes “fake news” to judicial review. French law, for example, stipulates that whether an item is “fake news,” and thus subject to removal or containment, is up to the determination of a judge. The ruling shall be made according to three criteria: the fake news must be manifest, disseminated deliberately on a massive scale, and lead to a disturbance of the peace or compromise the outcome of an election.¹⁶ The proportionate application of such a law is dependent on an independent judiciary insulated from political pressure, well-trained judges capable of understanding the digital information ecosystem, and a well-resourced judiciary capable of expediting the review of such claims, including any appeals.

Lawmakers and regulators should consider the range of approaches outlined in this text before resorting to the blunt-force instrument of a ban on or criminalization of fake news or disinformation. In instances where content and speech circulated on social media run afoul of existing criminal law, the referral of violating content for investigation and prosecution under such existing provisions – such as those covering defamation, hate speech, fraud, or identity theft -- is recommended over the adoption of additional criminal sanctions for the dissemination fake news or disinformation.

LEGAL AND REGULATORY

RESPONSES TO DISINFORMATION

2. MEASURES TO RESTRICT ONLINE CONTENT AND BEHAVIORS (/TOPICS/LEGAL/2-MEASURES-RESTRICT-ONLINE-CONTENT-AND-BEHAVIORS)

Measures to restrict content or behaviors related to the use of social media or other digital technologies strive to bring campaign regulations up to date with the current information environment. In the absence of rules of the road, social media and digital technologies can be used in campaigns in blatantly deceptive and destructive ways with impunity. While not explicitly prohibited, some uses of social media and other digital technologies may contradict principles governing campaigning enshrined elsewhere in the electoral law.

I. RESTRICT CONTENT OR BEHAVIORS: MEASURES DIRECTED AT DOMESTIC ACTORS

A. PROHIBIT SOCIAL MEDIA CAMPAIGNING OUTSIDE OF A DESIGNATED CAMPAIGN PERIOD

Many countries delimit the timeframe of the campaign period. This may consist of, for example, a stipulation that campaign activities may only begin one or several months before Election Day. An electoral silence period of one or several days directly prior to Election Day during which certain campaign activities are prohibited also has wide global precedent. These provisions may apply very narrowly to candidates and political parties contesting the election or more broadly to political statements or advertisements placed by non-contestant campaigners, meaning third parties engaged in campaigning that are not themselves candidates or political parties. Some countries have extended these provisions to consider the political activity and advertising on social media, but many are either silent on the topic of social media or explicitly exempt from campaign regulations.

Temporal restrictions on campaigning via social media are more likely to make an impact on the spread of disinformation when they are one part of a combination of measures intended to create rules and norms for the use of social media in campaigns. While disinformation tactics will continue to evolve, features of current online influence operations include the cultivation of online audiences, infiltration of existing online affinity networks, and the creation and growth of networks of coordinated accounts – processes that take time and, frequently, the investment of financial resources. Measures to temporally restrict the length of a campaign period combined

with detailed stipulations about what activities constitute a campaign expenditure, for example, might inhibit domestic actors seeking to build a deceptive social media presence over the course of months or years that they plan to activate during the campaign period.

Extending existing laws that set time restrictions on campaign periods to also cover social media can be relatively straightforward. Argentina's electoral laws, for example, indicate that television and radio advertising is limited to 35 days prior to the date set for the election and that campaigning via the internet or mobile technologies is only permissible during the campaign period (which starts 50 days before the Election Day and ends with the start of the electoral silence period 48 hours before elections).¹⁷ Resolution of the definitional considerations outlined in the section above – what constitutes digital media, online campaigning, and political advertising – is necessary to make the enforcement of restrictions on campaigning outside of the designated period predictable and proportionate.

Unlike some of the newer or more hypothetical legal and regulatory approaches explored elsewhere in this section of the guidebook, the interpretation of prohibitions on social media use during campaign periods has significant judicial precedent. Notable cases include:¹⁸

- In 2015 (<https://electionjudgments.org/en/entity/tm4sjrq6daq>), the High Chamber of the Federal Electoral Tribunal of Mexico ruled against a political party after a number of high-profile individuals tweeted in support of the party during the electoral silence period. The Tribunal determined that the coordination behind these actions, including the identification of paid intermediaries, constituted a part of the party's propaganda strategy.
- A 2010 ([https://electionjudgments.org/library/?q=\(allAggregations:!f,filters:\(%275bfb1a0471dd0fc16ada146%27\),unpublished:!f\)](https://electionjudgments.org/library/?q=(allAggregations:!f,filters:(%275bfb1a0471dd0fc16ada146%27),unpublished:!f))) ruling by the Superior Electoral Court of Brazil addresses an instance in which a Vice Presidential candidate tweeted in support of his Presidential running mate prior to the start of the campaign period. The court fined the candidate on the grounds that the tweet constituted illegal electoral propaganda.
- In two cases from 2012 (<https://electionjudgments.org/en/entity/4xu3lhyt9su>) and 2016 (<https://electionjudgments.org/en/entity/r19plihgo9c>), the High Chamber of the Federal Electoral Tribunal of Mexico ruled that candidates or pre-candidates posting to personal social media accounts outside of the campaign period were allowable if the content refrained from overt appeals for electoral support and was in the interest of free expression on issues of national interest.
- The Supreme Court of Slovenia (<https://www.rtvsl.si/news-in-english/supreme-court-on-election-blackouts-every-comment-is-not-propaganda/403791>) determined in 2016 that it was allowable to publish personal opinions during the electoral silence period, including via social media. The determination was made after a private citizen was fined for posting an interview with a candidate on Facebook during the silence period.

For regulators considering these measures, it should be noted that restrictions on the activities of legitimate political actors can provide an advantage to malign actors that are not subject to domestic law. Ahead of the 2017 French presidential elections, for example, [troves of hacked data](https://www.npr.org/sections/thetwo-way/2017/05/06/527154146/french-candidate-emmanuel-) (<https://www.npr.org/sections/thetwo-way/2017/05/06/527154146/french-candidate-emmanuel->

macron-says-campaign-has-been-hacked-hours-before-elec) from the campaign of Emmanuel Macron was posted online moments before the designated 24-hour silence period before Election Day during which media and campaigns are unable to discuss the election, leaving the campaign unable to respond publicly to the attack.

B. RESTRICT ONLINE BEHAVIORS THAT CONSTITUTE AN ABUSE OF STATE RESOURCES

The extension of Abuse of State Resources (ASR) provisions to social media is a way in which regulation (paired with enforcement) can deter incumbents from using the resources of the state to spread disinformation for political advantage. As domestic actors increasingly adopt tactics pioneered by foreign state actors to manufacture and artificially amplify social media content in deceptive ways to buoy their domestic political prospects, tactics to deter domestic corruption may have application.

The IFES ASR assessment framework

(https://www.ifes.org/sites/default/files/abuse_of_state_resources_research_and_assessment_framework.pdf) recognizes *Restrictions on Official Government Communications to the Public* and *Restrictions on State Personnel* as two elements of a comprehensive ASR legal framework for elections. These are two clear areas where extending ASR provisions to social media has value. For example, restrictions to the messages that an incumbent candidate might disseminate via public media may be logically extended to restrictions on the use of official government social media accounts for campaigning. Additionally, restrictions on state personnel – for example, banning engagement in campaigns while on duty or an overarching mandate to maintain impartiality – may need to be explicitly updated to address the use of personal social media accounts.

In regard to ASR, potential questions to be investigated could include – how are the official social media accounts of government agencies being used during the campaign period? Are the accounts of government agencies engaged in coordination with partisan social media accounts to promote certain narratives? How are the accounts of state employees being used to promote political content?

For incumbents seeking to use state resources to secure electoral advantage, the personal social media accounts of state personnel and the social media reach of official state agencies are attractive real estate in the mobilization of political narratives. In Serbia, for example, [analysis by the Balkan Investigative Reporting Network](https://balkaninsight.com/2020/06/18/the-castle-how-serbias-rulers-manipulate-minds-and-the-people-pay/) (<https://balkaninsight.com/2020/06/18/the-castle-how-serbias-rulers-manipulate-minds-and-the-people-pay/>) alleges that the ruling party maintained a software system that logged the actions of hundreds of individuals' social media accounts (many of those accounts belonging to state employees posting content during regular business hours) as they pushed party propaganda and disparaged political opponents ahead of 2020 elections. If true, these allegations would amount to a ruling party turning state employees into a troll army to wield against political opponents.

Prior to the 2020 elections, the Anti-Corruption Agency of Serbia issued a statement that “Political subjects and bearers of public functions should responsibly use social networks and the Internet for the pre-election campaign since political promotion on Internet pages owned by the government bodies represents an abuse of public resources.”¹⁹ The agency noted that the increase in campaigning via social media as a result of COVID-19 social distancing restrictions brought particular attention to this issue.

“The hardest thing is connecting bad actors to the government.... It’s not a grassroots problem; it’s an elite politics problem.” — Southeast Asian Civil Society Representative

Other actions that have been taken at the intersection of ASR and social media globally include a ruling by the Monterrey Regional Chamber of the Federal Electoral Tribunal of Mexico in 2015 (<https://electionjudgments.org/en/entity/bjb7r7wbb2>), which determined that in using a government vehicle to travel to polling stations with political candidates and posting about this activity via a Twitter account promoted on an official government webpage, a sitting governor violated the law. As a result, The court annulled the election, though a remedy this extreme is out of step with international good practice (https://www.ifes.org/sites/default/files/2018_ifes_when_are_elections_good_enough_final.pdf) on when elections can or should be annulled.

C. SET LIMITS ON THE USE OF PERSONAL DATA BY CAMPAIGNS

Restrictions on the use of personal data by domestic political actors are one avenue some countries are exploring to block the dissemination and amplification of disinformation. Microtargeting, the use of user data to precisely target advertisements and messages to highly specific audiences, has received considerable attention. Microtargeting may enable legitimate political entities, as well as malign foreign and domestic actors, to narrowly tailor advertising to reach highly specific audiences in ways that can enable the opaque dissemination of misleading or otherwise problematic content. By limiting campaigns’ ability to use personal data, regulators may also limit their ability to divisively target advertisements to very narrow audiences.

In the United Kingdom, the UK Information Commissioner’s Office (ICO) launched an investigation (<https://www.forbes.com/sites/kateoflahertyuk/2018/11/06/ico-acts-to-stop-data-misuse-in-elections/#2e07589b1ae6>) in 2017 to look at the use of personal data for political purposes in response to allegations that an individual’s personal data was being used to micro-target political advertisements during the EU Referendum. The ICO fined the Leave.EU campaign and associated entities for improper data protection practices and investigated the Remain campaign on a similar basis.

While the use of data is included in this section on restricting content or behaviors, the topic also has transparency and equity implications. In their [analysis of the regulation of online political microtargeting](https://policyreview.info/articles/analysis/regulation-online-political-micro-targeting-europe) (<https://policyreview.info/articles/analysis/regulation-online-political-micro-targeting-europe>) in Europe, academic Tom Dobber and colleagues note that a new Political Parties Act has been proposed in the Netherlands which “include[s] new transparency obligations for political parties with regard to digital political campaigns and political micro-targeting.”²⁰ Dobber goes on to observe that “The costs of micro-targeting and the power of digital intermediaries are among the main risks to political parties. The costs of micro-targeting may give an unfair advantage to the larger and better-funded parties over the smaller parties. This unfair advantage worsens the inequality between rich and poor political parties and restrains the free flow of political ideas.”²¹

Limitations on the use of personal data for political campaigning are generally included in larger policy debates around data privacy and individual’s rights over their personal data. In Europe, for example, the EU’s General Data Protection Regulation (GDPR) places restrictions on political parties’ ability to buy personal data, and voter registration records are inaccessible in most countries.²² The subject of data privacy is explored further in the topical section on [norms and standards](https://counteringdisinformation.org/node/2743/). (<https://counteringdisinformation.org/node/2743/>)

D. LIMIT POLITICAL ADVERTISING TO ENTITIES THAT ARE REGISTERED FOR THE ELECTION

Some jurisdictions limit the type of entities that are able to run political advertisements. Albanian electoral law, for example, stipulates that “only those electoral subjects registered for elections are entitled to broadcast political advertisements during the electoral period on private radio, television or audio-visual media, be they digital, cable, analog, satellite or any other form or method of signal transmission.”²³ In [Bowman v. the United Kingdom](https://hudoc.echr.coe.int/fre#%7B%22itemid%22:%5B%22001-58134%22%5D%7D) (<https://hudoc.echr.coe.int/fre#%7B%22itemid%22:%5B%22001-58134%22%5D%7D>), the European Court of Human Rights ruled that it is acceptable for countries to place financial limitations on non-contestant campaigning that is in line with limits for contestants, though the court also ruled that unduly low spending limits on non-contestants create barriers to their ability to freely share political views, violating Article 10 of the Convention.²⁴

Though candidates and parties may engage to various degrees in the dissemination of falsehoods and propaganda via their official campaigns, efforts intended to impact the information environment at scale will utilize unofficial accounts or networks of accounts to achieve their aims. Furthermore, such accounts are easily set up, controlled, or disguised to appear as though they are coming from extraterritorial locations, rendering national enforcement toothless.

In practice, measures to restrict advertisements run by a non-contestant would only be enforceable with compliance from social media companies – either through blanket restrictions on or pre-certification for political advertisements upheld by the platforms. Outside of a large market such as India or Indonesia, which have gained a degree of compliance from the platforms in enforcing such restrictions, this seems unlikely. The other route with the potential to make such

a measure enforceable would be if the platforms complied with government user-data requests from national oversight bodies that would seek to enforce violations. This presents a host of concerns for selective enforcement and potential violation of user privacy, particularly in authoritarian environments where such data could be misused to target opponents or other dissidents.

E. BAN THE DISTRIBUTION OR CREATION OF DEEPAKES FOR POLITICAL PURPOSES

Another legislative approach is to ban the use of deepfakes for political purposes. Several U.S. States have passed or proposed legislation to this effect, including [Texas](https://capitol.texas.gov/tlodocs/86R/billtext/html/SB00751S.htm) (<https://capitol.texas.gov/tlodocs/86R/billtext/html/SB00751S.htm>), [California](https://www.lexology.com/library/detail.aspx?g=4700f977-4845-417b-834d-b3c06390ee27) (<https://www.lexology.com/library/detail.aspx?g=4700f977-4845-417b-834d-b3c06390ee27>), and [Massachusetts](https://malegislature.gov/Bills/191/H3366.Html) (<https://malegislature.gov/Bills/191/H3366.Html>). Updates to [U.S. federal law](https://www.jdsupra.com/legalnews/first-federal-legislation-on-deepfakes-42346/) (<https://www.jdsupra.com/legalnews/first-federal-legislation-on-deepfakes-42346/>) in 2020 also require, among other things, the notification of the U.S. legislature by the executive branch in instances where foreign deepfake disinformation activities target US elections. Definitions of deepfakes in these pieces of legislation focus on an intent to deceive through highly realistic manipulation of audio or video using artificial intelligence.

It is conceivable that existing statutes related to identifying fraud, defamation, or consumer protection might cover the deceptive use of doctored videos and images for political purposes. [One study](https://sensity.ai/mapping-the-deepfake-landscape/) (<https://sensity.ai/mapping-the-deepfake-landscape/>) reports that 96 percent of deepfakes involve the nonconsensual use of female celebrities' images in pornography, suggesting that existing provisions related to identity fraud or non-consensual use of intimate imagery may also be applicable. Deepfakes are often used to discredit women candidates and public officials, so sanctioning the creation and/or distribution of deepfakes, or using existing legal provisions to prosecute the perpetrators of such acts, could have an impact on disinformation targeting women that serve in a public capacity.

F. CRIMINALIZE DISSEMINATION OF FAKE NEWS OR DISINFORMATION

One common approach to regulation is the introduction of legal provisions that criminalize the disseminators or creators of disinformation or fake news. This is a worrisome trend as it has significant implications for freedom of expression and freedom of the press. As discussed in the definition section [Why are definitions of fake news and disinformation problematic?](https://counteringdisinformation.org/topics/legal/1-definitions#FakeNewsProblematic) (<https://counteringdisinformation.org/topics/legal/1-definitions#FakeNewsProblematic>), the extreme difficulty of arriving at clear definitions of prohibited behaviors can lead to unjustified restrictions and direct harms to human rights. Though some countries adopt such measures in recognition of and out of an attempt to mitigate the impact of disinformation on political and electoral processes, such provisions are also opportunistically adopted by regimes to stifle political opposition and muzzle the press. Even in countries where measures might be undertaken in a good faith attempt to protect democratic spaces, the potential for abuse and selective

enforcement is significant. Governments have also passed a number of restrictive and emergency laws in the name of curbing COVID-related misinformation and disinformation (<https://www.ifes.org/publications/ifes-covid-19-briefing-series-preserving-electoral-integrity-during-infodemic>) with similarly chilling implications for fundamental freedoms. The Poynter Institute maintains a database of anti-misinformation laws (<https://www.poynter.org/ifcn/anti-misinformation-actions/#kenya>) with an analysis of their implications.

Before adopting additional criminal penalties for the dissemination of disinformation, legislators and regulators should consider whether existing provisions in the criminal law such as those covering defamation, hate speech, identity theft, consumer protection, or the abuse of state resources are sufficient to address the harms that new criminal provisions attempt to address. If the existing criminal law framework is deemed insufficient, revisions to criminal law should be undertaken with caution and awareness of the potential for democratically damaging downstream results.

"If we want to fight hoaxes, it's not through criminal law, which is too rigid." — Indonesian Civil Society Representative

It should be noted that some attempts have been made to legislate against online gender-based violence, which sometimes falls into the category of disinformation. Scholars Kim Barker and Olga Jurasz consider this question in their book, Online Misogyny as Hate Crime: A Challenge for Legal Regulation? (<https://www.routledge.com/Online-Misogyny-as-Hate-Crime-A-Challenge-for-Legal-Regulation-1st-Edition/Barker-Jurasz/p/book/9781138590373>), where they conclude that existing legal frameworks have been unsuccessful in ending online abuse because they focus more on punishment after a crime is committed rather than on prevention.

II. RESTRICT CONTENT OR BEHAVIORS: MEASURES DIRECTED AT SOCIAL MEDIA AND TECHNOLOGY PLATFORMS

National legislation directed at social media and technology platforms is often undertaken in an attempt to increase domestic oversight over these powerful international actors who have little legal obligation to minimize the harms that stem from their products. Restrictions on content and behaviors that compel platform compliance can make companies liable for all of the content on their platforms, or more narrowly target only the paid advertising on their platforms. In this debate, platforms will argue, with some merit, that it is nearly impossible for them to screen billions of daily individual user posts. Conversely, it may be more reasonable to expect social media platforms to scrutinize *paid* advertising content.

As discussed in the section on domestic actors, some countries prohibit paid political advertising outside of the campaign period, some restrict paid political advertising altogether, while others limit the ability to place political advertisements only to entities that are registered for the election. In some instances, countries have called on social media companies to enforce these restrictions by making them liable for political advertisements on their platforms.

Placing responsibility on the platforms to enforce national advertising restrictions also has the potential to create a barrier for political or issue advertisements placed by seemingly non-political actors or by unofficial accounts affiliated with political actors. However, if national regulators do take this approach, the difficulties of compliance with dozens if not hundreds of disparate national regulatory requirements are certain to be a point of contention with the companies. Like any other measure that places boundaries on permissible political expression, it also carries the potential for abuse.

The global conversation around platform regulations that would fundamentally alter the business practices of social media and technology companies – anti-trust or user data regimes, for example – are beyond the scope of this chapter. The focus instead is attempts at a national level to place enforceable obligations on the platforms that alter the way that they conduct themselves in a specific national jurisdiction.

Often, the enforceability of country-specific regulations placed on the platforms will differ based on the perceived political or reputational risk associated with inaction in a country, which can be associated with market size, geopolitical significance, potential for electoral violence, or international visibility. That being said, some measures are more easily complied with in the sense that they do not require platforms to reconfigure their products in ways that have global ramifications and thus are more easily subject to national rule making.

The ability of a country to compel action from the platforms can also be associated with whether the platforms have an office or legal presence in that country. This reality has spawned national laws requiring platforms to establish a local presence to respond to court orders and administrative proceedings. [Germany](https://www.insidetechmedia.com/2020/02/13/germany-likely-to-adopt-unique-regulatory-regime-for-intermediaries-to-media-services/) (<https://www.insidetechmedia.com/2020/02/13/germany-likely-to-adopt-unique-regulatory-regime-for-intermediaries-to-media-services/>) has included a provision to this effect in their Interstate Media Treaty. Requirements to appoint local representatives that enable platforms to be sued in court become highly contentious in countries that lack adequate legal protections for user speech and where fears of censorship are well-founded. A controversial [Turkish law](https://www.ft.com/content/91c0a408-6c15-45c3-80e3-d6b2cf913070) (<https://www.ft.com/content/91c0a408-6c15-45c3-80e3-d6b2cf913070>) went into effect on October 1, 2020 requiring companies to appoint a local representative accountable to local authorities' orders to block content deemed offensive. U.S.-based social media companies have [chosen not to comply](https://www.ft.com/content/91c0a408-6c15-45c3-80e3-d6b2cf913070) (<https://www.ft.com/content/91c0a408-6c15-45c3-80e3-d6b2cf913070>) at the urging of human rights groups, and [face escalating fines and possible bandwidth restrictions](https://www.hurriyetdailynews.com/turkey-fines-facebook-and-possible-bandwidth-restrictions) ([https://www.hurriyetdailynews.com/turkey-fines-facebook-](https://www.hurriyetdailynews.com/turkey-fines-facebook-and-possible-bandwidth-restrictions)

[others-over-new-social-media-law-159732](#)) that would throttle access to the platforms in Turkey in the case of continued non-compliance. This contrast illustrates that challenges social media platforms must navigate in complying with national law. Measures that constitute reasonable oversight in a country with robust protections for civil and political rights might serve as a mechanism for censorship in another.

At the same time, joint action grounded in international human rights norms could be one way for countries with less individual influence over the platforms to elevate their legitimate concerns. The Forum on Information and Democracy's [November 2020 Policy Framework](#) (https://informationdemocracy.org/wp-content/uploads/2020/11/ForumID_Report-on-infodemics_101120.pdf) articulates the challenge of harmonizing transparency requirements while preventing politically motivated abuse of national regulations. While joint action at the level of the European Union is occurring, the report points to the possibility of the Organization for American States, the African Union, the Asia-Pacific Economic Cooperation or Association of Southeast Asian Nations, or regional development banks as potential organizing forums for joint action in other regions.

A. HOLD PLATFORMS LIABLE FOR ALL CONTENT AND REQUIRE REMOVAL OF CONTENT

The debate over what content should be allowable on social media platforms is global in scope. Analysis on this topic is prolific and global consensus is unlikely to emerge given legitimate and differing definitions of the bounds that can and should be placed on speech and expression. Many of these measures that introduce liability for all content have hate speech as a central component. While hate speech is not limited to political or electoral periods, placing pressure on societal fault lines through the online [amplification of hate speech](#) (<https://www.ifes.org/publications/disinformation-campaigns-and-hate-speech-exploring-relationship-and-programming>) is a common tactic used in political propaganda and by disinformation actors during electoral periods.

Some national jurisdictions have attempted to introduce varying degrees of platform responsibility for all the content hosted on their platforms, regardless of whether that is organic or paid content.

The German Network Enforcement Act

([http://www.bundesrat.de/SharedDocs/drucksachen/2017/0501-0600/536-17.pdf?](http://www.bundesrat.de/SharedDocs/drucksachen/2017/0501-0600/536-17.pdf?__blob=publicationFile&v=1)

[__blob=publicationFile&v=1](#)) (NetzDG) requires social media companies to delete “manifestly unlawful” content within 24 hours of being notified. Other illegal content must be reviewed within seven days of being reported and deleted if found to be in violation of the law. Failure to comply

carries up to a 5 million euro fine, though the law exempts providers who have fewer than 2 million users registered in Germany. The law does not actually create new categories of illegal content; its purpose is to require social media platforms to enforce 22 statutes on online content that already exist in the German code. It targets already-unlawful content such as “public incitement to crime,” “violation of intimate privacy by taking photographs,” defamation, “treasonous forgery,” forming criminal or terrorist organizations, and “dissemination of depictions of violence.” It also includes Germany’s well-known prohibition of glorification of Nazism and Holocaust denial. The takedown process does not require a court order or provide a clear appeals mechanism, relying on online platforms to make these determinations.²⁵

The law has been criticized (<https://www.article19.org/wp-content/uploads/2017/12/170901-Legal-Analysis-German-NetzDG-Act.pdf>) as being too broad and vague in its differentiation of “unlawful content” and “manifestly unlawful content.” Some critics also object (https://www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf) to NetzDG as a “privatized enforcement” law because online platforms assess the legality of the content, rather than courts or other democratically legitimate institutions. It is also credited with inspiring a number of copycat laws in countries where the potential for censoring legitimate expression is high. As of late 2019, *Foreign Policy* (<https://foreignpolicy.com/2019/11/06/germany-online-crackdowns-inspired-the-worlds-dictators-russia-venezuela-india/>) identified 13 countries that had introduced similar laws; the majority of these countries were ranked as “not free” or “partly free” in Freedom House’s 2019 Freedom of the Internet assessment.²⁶

France, which has pre-existing rules restricting hate speech, also introduced measures (<https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000042031970>) similar to those in Germany to govern content online. However, the French constitutional court overturned (<https://www.nytimes.com/2020/06/18/world/europe/france-internet-hate-speech-regulation.html>) these measures in 2020, which similar to the German law would have required platforms to review and remove hateful content flagged by users within 24 hours or face fines. The court ruled that the provisions in the law would lead platforms to adopt an overly conservative attitude toward removing content in order to avoid fines, thus restricting legitimate expression.

The United Kingdom is another frequently cited example that illustrates various approaches to regulating harmful online content, including disinformation. A 2019 Online Harms White Paper (<https://www.gov.uk/government/consultations/online-harms-white-paper>) outlining the UK government’s plan for online safety proposed placing a statutory duty of care on internet companies for the protection of their users, with oversight by an independent regulator. A public consultation period for the Online Harms Paper informed proposed legislation in 2020 that focuses on making the companies responsible (<https://www.reuters.com/article/us-britain-tech-regulation/uk-to-make-social-media-platforms-responsible-for-harmful-content-idUSKBN2060Q7>) for the systems they have in place to protect users from harmful content. Rather than require companies to remove specific pieces of content, the new framework would require the platforms to provide clear policies on the content and behavior that are acceptable on their sites and enforce these standards consistently and transparently.

These approaches contrast with the Bulgarian framework, for example, which exempts social media platforms from editorial responsibility.²⁷ Section 230 of the Communications Decency Act (<https://www.law.cornell.edu/uscode/text/47/230>) of the United States law also expressly releases social media platforms from vicarious liability.

Other laws have been proposed or enacted in countries around the globe that introduce some degree of liability or responsibility for platforms to moderate harmful content on their platforms. Broadly speaking, this category of regulatory response is the subject of fierce debate on the potential for censorship and abuse. The models in Germany, France, and the United Kingdom have frequently cited examples of attempts by consolidated democracies to more actively imposing a duty on platforms for the content they host while incorporating sufficient checks to protect freedom of expression – though measures in all three countries are also criticized for the ways they have attempted to strike this balance. These different approaches also illustrate how a proliferation of national laws introducing platform liability is poised to place a multitude of potentially contradictory obligations on social media companies.

B. PROHIBIT PLATFORMS FROM HOSTING PAID POLITICAL ADVERTISING

Some jurisdictions prohibit paid campaign advertising in traditional media outright, with that ban extending or potentially extending to paid advertising on social media.²⁸ “For decades, paid political advertising on television has been completely banned during elections in many European democracies. These political advertising bans aim to prevent the distortion of the democratic process by financially powerful interests and to ensure a level playing field during elections.”²⁹

The French Electoral Code (https://www.cjoint.com/doc/20_01/JAhm1cW3lBh_codeelectoral.pdf) stipulates that for the 6 months prior to the month of an election, commercial advertising for the purposes of election propaganda via the press or “any means of audiovisual communication” is prohibited.³⁰ A stipulation such as this is contingent on clear definitions of online campaigning and political advertising; amendments (<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000037847559&categorieLien=id>) to the French Electoral Code in 2018, for example, attempt to inhibit a broad range of political and issue advertisements by stipulating that the law applies to “information content relating to a debate of general interest,”³¹ rather than limiting the provision to ads that directly reference candidates, parties, or elections. In the French case, these provisions along with a number of transparency requirements discussed in the sections below, led some platforms, such as Twitter, to ban all political campaign ads and issue advocacy ads in France, a move that was later expanded into a global policy (<https://business.twitter.com/en/help/ads-policies/ads-content-policies/political-content.html>). Similarly, Microsoft banned all ads in France “containing content related to debate of general interest linked to an electoral campaign,” which is also now a global policy (<https://about.ads.microsoft.com/en-us/resources/policies/disallowed-content-policies>). Google banned all ads containing “informational content relating to a debate of general interest” between April and May 2019 across its platform in France, including YouTube.³² The French law led Twitter to initially block (<https://apnews.com/article/d0e60c2130064a7f87469113382b7001>) an attempt by

the French Government's information service to pay for sponsored tweets for a voter registration campaign in the lead-up to European parliamentary elections, though this position was eventually reversed.

The French ban on issue advertising on social media was legitimated by a parallel ban on political advertising via print or broadcast media. Other jurisdictions seeking to impose restrictions on social media advertising might similarly consider aligning those rules with the principles governing offline or traditional media advertising.

C. HOLD PLATFORMS RESPONSIBLE FOR ENFORCING RESTRICTIONS ON POLITICAL ADVERTISEMENTS RUN OUTSIDE A DESIGNATED CAMPAIGN PERIOD

Some jurisdictions have opted to place responsibility on the entities that sell political advertisements, including social media companies, to enforce restrictions on advertising outside of the designated campaign period – both before the campaign period begins as well as during official silence periods in the day or days directly before the election.

Indonesia had some success calling on the platforms to enforce the three-day blackout period prior to its 2019 Elections. According to interlocutors, Bawaslu sent a letter to all of the platforms advising them that they would enforce criminal penalties should the platforms allow paid political advertising on their platforms during the designated blackout period. Despite responses from one or more of the platforms that the line between advertising in general and political advertising was too uncertain to enforce a strict ban, Bawaslu insisted that the platforms find a way to comply. The platforms in turn reported rejecting large numbers of advertisements during the blackout period. Bawaslu's restrictions applied only to paid advertising, not organic posts.

Under India's "[Voluntary Code of Ethics for the 2019 General Election](https://eci.gov.in/files/file/9468-voluntary-code-of-ethics-by-the-social-media-platforms-for-the-general-election-2019/) (<https://eci.gov.in/files/file/9468-voluntary-code-of-ethics-by-the-social-media-platforms-for-the-general-election-2019/>)," social media companies committed themselves to take down prohibited content within three hours during the 48-hour silence period before polling. The signatories to the Code of Ethics developed a notification mechanism through which the Election Commission could inform relevant platforms of potential violations of Section 126 of the Representation of the People Act, which bars political parties from advertising or broadcasting speeches or rallies during the silence period.

India and Indonesia are both very large markets, and most global social media companies have a physical presence in both countries. These factors significantly contribute to these countries' abilities to compel platform compliance. This route is unlikely to be as effective in countries that do not have as credible a threat of legal sanction over the platforms or the ability to place penalties or restrictions on the platforms in a way that impacts their global business.

For countries that do attempt this route, as with restrictions on social media campaigning placed on domestic actors, restrictions that rely on the platforms for enforcement must also acknowledge the definitional distinctions between paid and unpaid content and between political

and issue campaigning, for example, to have any enforceability. The Canadian framework acknowledges the complexity of enforcing campaign silence online by exempting content that was in place before the blackout period and has not been changed.³³ Facebook's decision (<https://www.nytimes.com/2020/09/03/technology/facebook-election-chaos-november.html>) to unilaterally institute a political advertising blackout period for the time period directly surrounding the 2020 U.S. Elections also limited political advertising to content already running on the platform. No ads containing new content could be placed. Moves to restrict paid advertising may advantage incumbents or other contestants that have had time to establish a social media audience in advance of the election; paid advertising is a critical tool that can allow new candidates to reach large audiences.

D. ONLY ALLOW PLATFORMS TO RUN PRE-CERTIFIED POLITICAL ADVERTISEMENTS

During the 2019 elections, the Election Commission of India required that paid online advertising that featured the names of political parties or candidates be vetted and pre-certified by the Election Commission. Platforms, in turn, were only allowed to run political advertisements that had been pre-certified.³⁴

This measure only applied to a narrow band of political advertisements – any issue ads or third-party ads that avoid explicit mention of parties and candidates would not need to be pre-certified under these rules. For other countries, implementation of a pre-certification requirement would necessitate institutional capacity on par with Indian electoral authorities to make the vetting of all ads possible, as well as the market size and physical presence of company offices in-country to get the companies to comply.

Mongolia's draft electoral laws would require political parties and candidates to register their websites and social media accounts. These draft laws would also block access to websites that run content by political actors that do not comply. The provision worded as such seems to penalize third-party websites for breaches committed by a contestant. Provisions further require that the comments function on official campaign websites and social media accounts should be disabled, and non-compliance with this provision incurs a fine.³⁵ As the law is still in draft form, the enforceability of these measures has not been tested at the time of publication.

E. OBLIGATE PLATFORMS TO BAN ADVERTISEMENTS PLACED BY STATE-LINKED MEDIA

At present, social media platforms have differing policies on the ability of state-controlled news media to place paid advertising on their platforms. While platforms have largely adopted restrictions on foreign actors' ability to place *political advertising*, some platforms still allow state-controlled media to pay to promote their content to foreign audiences more generally. Twitter has banned state-controlled media entities from placing paid advertising of any kind on their platform.³⁶ For countries where Facebook's Ad Library is being enforced, the advertiser verification process (<https://www.facebook.com/business/help/2150157295276323>) attempts to

prohibit foreign actors from placing political advertising. However, Facebook does not currently restrict the ability of state-linked media to pay to promote their news content to foreign audiences, a tool that state actors use to build foreign audiences.

Analysis by the Stanford Internet Observatory (<https://cyber.fsi.stanford.edu/news/chinese-state-media-shapes-coronavirus-convo>) demonstrates how Chinese state media uses social media advertising as a part of broader propaganda efforts and how such efforts were used to build a foreign audience for state-controlled traditional media outlets and social media accounts. The ability to reach this large audience was then used to deceptively shape favorable narratives about China during the coronavirus pandemic.

Prohibitions against foreign state-linked actors paying to promote their content to domestic audiences could be tied to other measures that attempt to bring transparency in political lobbying. For example, some experts (<https://www.washingtonpost.com/opinions/2020/04/27/chinese-propaganda-covid-19-grows-us-social-media-must-act/>) in the U.S. propose applying the Foreign Agents Registration Act (FARA) to restrict the ability of foreign agents registered under FARA to advertise to American audiences on social media. This in turn requires a consistent and proactive effort on the part of U.S. authorities to require that state media is identified and registered as foreign agents. Rather than prohibit ads placed by known foreign agents, another option is to require platforms to label such ads to increase transparency. Several platforms have independently adopted such provisions,³⁷ though enforcement has been inconsistent (<https://www.propublica.org/article/youtube-promised-to-label-state-sponsored-videos-but-doesnt-always-do-so>).

F. RESTRICT HOW PLATFORMS CAN TARGET ADVERTISEMENTS OR USE PERSONAL DATA

Another avenue being explored in larger markets is placing restrictions on the ways in which personal data can be used by platforms to target advertising. Platforms, to some degree, are adopting such measures in the absence of specific regulation. Google, for example, allows a narrower range of targeting criteria (<https://support.google.com/adspolicy/answer/6014595?hl=en>) to be used to place election ads compared to other types of advertisements. Facebook does not limit the targeting of political ads, though they offer various tools (<https://about.fb.com/news/2020/01/political-ads/>) to provide a degree of transparency for users on how they are being targeted. Facebook also allows users to opt-out of certain political ads (<https://about.fb.com/news/2020/06/voting-information-center/>), though these options are only available in the United States as of early 2021. Less well-understood are the tools used by streaming television services (<https://www.cnn.com/2020/06/03/politics/streaming-services-political-ads/index.html>) to target ads. It is unlikely that national-level regulation of this nature outside of the U.S. or EU will have the ability to alter the platforms' policies. Further discussion on this topic can be found in the topical section on platform responses to disinformation (<https://counteringdisinformation.org/node/2722/>).

LEGAL AND REGULATORY RESPONSES TO DISINFORMATION

3. MEASURES TO PROMOTE TRANSPARENCY DURING CAMPAIGNING AND ELECTIONS (/TOPICS/LEGAL/3-MEASURES-PROMOTE-TRANSPARENCY-DURING-CAMPAIGNING-AND-ELECTIONS)

Measures that **promote transparency** can include obligations for domestic actors to disclose the designated political activities they engage in on social media, as well as obligations for digital platforms to disclose information on the designated political activities that take place on their platforms or to label certain types of content that may otherwise be misleading. These measures are part of the regulatory push back against disinformation as they allow insight into potentially problematic practices being used by domestic political or foreign actors and build public understanding of the origins of the content they are consuming. Transparency creates the opportunity for the public to make better-informed decisions about their political information.

I. PROMOTE TRANSPARENCY: MEASURES DIRECTED AT DOMESTIC ACTORS

A. REQUIRE THE DECLARATION OF SOCIAL MEDIA ADVERTISING AS A CAMPAIGN EXPENDITURE

One of the most common approaches to promoting increased transparency by domestic actors is to expand the definition of “media” or “advertising” that is subject to existing disclosure requirements to include online and social media advertising. Expansions of this nature should take into account the definitional considerations at the beginning of this section of the guidebook. Detailed disclosure requirements may be required to delineate which types of expenditures constitute social media advertisements, including, for example, payments to third parties to post supportive content or attack opponents. While expanding existing disclosure requirements extends existing principles of transparency, crafting meaningful disclosure requirements necessitates careful consideration of the ways in which social media and online advertising differ from non-digital forms of political advertising.

To offer illustrative examples, section 349 of [Canada's Elections Act](https://laws-lois.justice.gc.ca/PDF/E-2.01.pdf) (<https://laws-lois.justice.gc.ca/PDF/E-2.01.pdf>) has extensive regulation on third-party expenditure and the use of foreign funding, which captures paid advertising online. A [draft resolution in Colombia](https://drive.google.com/file/d/1Wh5vNUhZV5iQytTXeT0bHSx316LCYiSE/view) (<https://drive.google.com/file/d/1Wh5vNUhZV5iQytTXeT0bHSx316LCYiSE/view>) has also been put forth with the aim of categorizing paid advertising on social media as a campaign expenditure

subject to spending limits. The resolution would empower Colombian electoral authorities to investigate these expenditures, given that they are often incurred by third parties and not by the campaign itself. It would establish a register of online media platforms that sell political advertising space and subject political advertising on social media to the same framework as political campaigning in public spaces.

B. REQUIRE REGISTRATION OF PARTY AND CANDIDATE SOCIAL MEDIA ACCOUNTS

While monitoring the official accounts of parties and candidates provides only a narrow glimpse into political advertising and political messages circulating on social media, having a record of official social media accounts is a first step toward transparency. This could be achieved by requiring candidates and parties to declare the accounts that are administered by or financially linked to their campaigns. This approach can provide a starting point for oversight bodies to monitor compliance with local laws and regulations governing campaigning. Such a requirement could be paired with a regulation that stipulates that candidates and campaigns may only engage in certain campaign activities through registered social media accounts, such as paying to promote political content or issue ads. This combination of measures can create an avenue for enforcement in instances where parties or candidates are found to be using social media accounts in prohibited ways of concealing financial relationships with nominally independent accounts. Enforcement would necessitate monitoring for compliance, which is discussed in the Enforcement subsection (<https://counteringdisinformation.org/topics/legal/6-enforcement>) at the end of this topical section of the guidebook.

This approach has been taken in Tunisia, where a directive issued by the country's election commission requires candidates and parties to register their official social media accounts with the commission.³⁸ Mongolia's draft election laws would also impose an obligation for the candidate, party, and coalition websites and social media accounts to be registered with the Communications Regulatory Commission (for parliamentary and presidential elections) and with the respective election commission (for local elections).³⁹ The Mongolian law in its entirety should not, however, be taken as a model as it raises concerns related to freedom of expression and enforcement limitations given definitional vagueness.

C. REQUIRE DISCLOSURE AND LABELING OF BOTS OR AUTOMATED ACCOUNTS

"Bots" or "Social Bots," which can perform automated actions online that mimic human behaviors, have been used as a part of disinformation campaigns in the past, though the degree to which they have impacted electoral outcomes is disputed.⁴⁰ When deployed by malign actors in the information space, these lines of code can, for example, power artificial social media personas, generate and amplify social media content in large quantities, and be mobilized to harass legitimate social media users.

As public awareness of this tactic has grown, lawmakers have attempted to legislate in this area to mitigate the problem. Legislative approaches that seek to ban the use of bots have largely failed to gain traction. A [measure](http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_R1U8H0S1L3R1F1J7M0B2E2C3W1Y9Y9) to criminalize bots or software used for online manipulation was proposed in South Korea, for example, but ultimately was not enacted. A [proposed bill](https://www.oireachtas.ie/en/bills/bill/2017/150/) in Ireland to criminalize the use of a bot to post political content through multiple fake accounts also failed to become law.

Opinion is divided [on the efficacy and freedom of expression implications of such measures](https://www.newyorker.com/tech/annals-of-technology/will-californias-new-bot-law-strengthen-democracy). Detractors of this approach suggest that such legislation can inhibit political speech and that overly broad measures can undermine legitimate political uses for bots, such as a voter registration drive or an electoral authority using a chatbot to respond to common voter questions. Detractors also suggest that legislating against specific disinformation tactics is a losing battle given that tactics evolve so quickly. Removing networks of automated bots also aligns with social media platforms' reputational self-interest, so that legislation against such operations may not be necessary.

Efforts to add transparency and disclosure to the use of bots may be a less controversial approach than criminalizing their use. California [passed a law](https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1001) in 2019 making it illegal to “use a bot to communicate or interact with another person in California online with the intent to mislead the other person about its artificial identity.” Germany's Interstate Media Treaty (Medienstaatsvertrag – “MStV”) also includes provisions that promote transparency around bots by obligating platforms to [identify and label content that is disseminated by bots](https://www.insidetechnology.com/2020/02/13/germany-likely-to-adopt-unique-regulatory-regime-for-intermediaries-to-media-services/). Measures that criminalize or require disclosure of the use of bots do present challenges for enforcement given [difficulty in reliably identifying bots](https://michaelkreil.github.io/openbots/).

“By the time lawmakers get around to passing legislation to neutralize a harmful feature, adversaries will have left it behind.” — Renee DiResta
(<https://www.ribbonfarm.com/2018/11/28/the-digital-magnot-line/>), Research Director at the Stanford Internet Observatory

D. REQUIRE DISCLOSURE OF THE USE OF POLITICAL FUNDS
ABROAD

Facing tightening regulations in their home countries, political actors might also seek to place political advertisements on social media by coordinating with actors located outside of the country. Foreign funding might also be used to place advertisements that target diaspora communities eligible for out-of-country voting. While platforms with political ad disclosure and identification requirements will in some cases prohibit the purchase of political advertisements in foreign currencies or by accounts operated from another country, these efforts are not yet sufficient to catch all political or issue advertisements placed extraterritorially.

Disclosure requirements that address foreign funding may wish to consider the ways in which foreign expenditures on social media advertising might differ from traditional media. New Zealand, for example, requires full disclosure of any advertising purchased by entities outside of the country, so that non-abidance constitutes a campaign finance violation.⁴¹ It could, however, be difficult to prove the beneficiary political party or candidate is aware of campaign funding being expended to their benefit extraterritorially, which could render enforcement futile.

II. PROMOTE TRANSPARENCY: MEASURES DIRECTED AT PLATFORMS

A. REQUIRE PLATFORMS TO MAINTAIN AD TRANSPARENCY REPOSITORIES

Some countries have imposed legal obligations on larger online platforms to maintain repositories of the political advertisements purchased on their platforms. France and Canada, for instance, require large online platforms to maintain a political ad library. India's Code of Ethics, signed by social media companies operating in the country ahead of 2019 elections, committed signatories to "facilitating transparency in paid political advertisements, including utilizing their pre-existing labels/disclosure technology for such advertisements." This measure may have been decisive in compelling these companies to expand coverage of their ad transparency features to India.

Facebook voluntarily introduced a publicly accessible Ad Library in a very limited number of countries in 2018, and as of early 2021 has since expanded coverage to 95 countries and territories (<https://www.facebook.com/business/help/2150157295276323>). Google maintains political ad transparency (<https://transparencyreport.google.com/political-ads/home?hl=en>) disclosures for Australia, the EU and UK, India, Israel, New Zealand, Taiwan, and the United States but has been slower to expand these tools to additional markets. As platforms contemplate where to next expand their advertising transparency tools, it is conceivable that updating national law to require platforms to maintain ad repositories could influence how companies prioritize countries for expansion. Details on the functionality of advertising transparency tools can be found in the guidebook section covering platform responses to disinformation (<https://counteringdisinformation.org/node/2722/>).

Legal mandates, however, might disadvantage smaller online platforms, since the cost of setting up and maintaining advertising repositories might be disproportionately higher for smaller platforms than for larger platforms. The legal requirement might thereby inadvertently stifle

platform plurality and diversity. This side effect can be remedied by creating a user threshold for the obligation. For example, Canada's ad transparency requirements apply only to platforms with more than three million regular users in Canada,⁴² though even this threshold might be too low to avoid becoming a barrier to competition. National regulators might also consider a standard whereby a platform is required to provide ad transparency tools if a certain percentage of the country's population uses their services.

Some countries where the platforms do not maintain ad repositories have experimented with their own. Ahead of the 2019 elections, South Africa tested a new political ad repository, built in partnership with election authorities and maintained by civil society. Compliance was not obligatory and was accordingly minimal among political parties, but the effort showed sufficient promise that the implementers of the ad repository are considering making compliance legally mandatory for future elections.⁴³

Legal measures that compel, or attempt to compel, platforms to maintain ad repositories might also incorporate provisions requiring the clear labeling of advertisers to distinguish between paid and organic content, as well as labels that distinguish among advertisements, editorial, and news content. Requirements to label content originating from state-linked media sources might also be outlined. Measures might also include identity verification requirements for actors or organizations that run political and issue advertisements. However, these provisions would likely require alterations to the functionality of the platform's ad transparency tools, a change that is more likely with joint pressure from multiple countries.

B. REQUIRE PLATFORMS TO PROVIDE ALGORITHMIC TRANSPARENCY

Additional measures being explored in France, Germany, and elsewhere focus on compelling platforms to provide greater insight into the algorithms that influence how content – organic and paid – is surfaced to individual users, or, put another way, transparency for users into how their data is used to inform the ads and content that they see.

Germany's MStV law, for example, introduces new definitions and rules intended to promote transparency across a comprehensive array of online portals and platforms. "Under the transparency provisions, intermediaries will be required to provide information about how their algorithms operate, including: [1] The criteria that determine how content is accessed and found. [2] The central criteria that determine how content is aggregated, selected, presented and weighed."⁴⁴ EU law on comparable topics has in the past drawn on German law to inform its development, suggesting that this route may influence conversations at the EU-level on platform transparency and, subsequently, include the global operations of digital media providers and intermediaries.

The Forum on Information and Democracy's [November 2020 Policy Framework](https://informationdemocracy.org/wp-content/uploads/2020/11/ForumID_Report-on-infodemics_101120.pdf) (https://informationdemocracy.org/wp-content/uploads/2020/11/ForumID_Report-on-infodemics_101120.pdf) provides a detailed discussion on how algorithmic transparency might be

regulated by state actors.⁴⁵

LEGAL AND REGULATORY RESPONSES TO DISINFORMATION

4. MEASURES TO PROMOTE EQUITY DURING CAMPAIGNS AND ELECTIONS (/TOPICS/LEGAL/4-MEASURES- PROMOTE-EQUITY-DURING-CAMPAIGNS- AND-ELECTIONS)

Measures designed to **promote equity** can include creating and enforcing spending caps for political parties and candidates with the goal of creating a level playing field for less financially well-resourced contenders. Other countries are experimenting with obligations for platforms to provide equitable advertising rates or provide free, equitably available ad space to candidates and parties.

Promoting equity as a deterrent to disinformation is an acknowledgment of the financial foundations of many coordinated disinformation campaigns. By providing political contestants with more equitable opportunities to be heard by the electorate, these measures attempt to lessen the advantage of financially well-resourced contenders who may – among other tactics – direct resources toward the promotion of disinformation to skew the information space. Strategies that promote equity can also benefit women, people with disabilities, and people from marginalized groups who are often less well-resourced than their more privileged counterparts and who are often targets of disinformation campaigns.

I. PROMOTE EQUITY: MEASURES DIRECTED AT DOMESTIC ACTORS

A. CAP PARTY OR CANDIDATE SOCIAL MEDIA EXPENDITURES

An approach to leveling the playing field on social media is capping how much each party or candidate can spend on social media, either as an absolute cap or as a percentage of overall campaign spending.

Romania, for example, caps expenditure for paid social media advertising at 30 percent of the overall allowed spending.⁴⁶ In the U.K., spending on social media is counted toward candidates' and parties' applicable spending limit and must be reported. Any material published on social

media that is “election material” – i.e., promotes or opposes: specific political parties, candidates or parties that support particular policies or issues, or types for candidates, and is made available to the public – counts toward the limit.⁴⁷

These measures do, however, require respective countries to operate effective campaign spending disclosure and investigation mechanisms—an asset most democracies lack.

II. PROMOTE EQUITY: MEASURES DIRECTED AT PLATFORMS

A. REQUIRE PLATFORMS TO PUBLISH ADVERTISING RATES AND TREAT ELECTORAL CONTESTANTS EQUALLY

Multiple countries have updated their legal frameworks to extend the principle of equity in the pricing of political advertisements to social media. In the context of traditional media, legal and regulatory measures might be used to ensure that candidates and parties have access to the same advertising opportunities at the same price. For example, measures requiring television, radio, or print media to publish their advertising rates as a means to ensure all actors have equal access to these distribution channels and that outlets cannot censor certain political views by charging different rates.

Extending this logic to social media – where advertising views are often determined in real-time online auctions that take place in the blink of an eye as users scroll through their social media feeds or refresh their internet browsers – presents a different challenge. The cost to place an ad will fluctuate based on numerous factors that determine how much demand exists to reach specific users. For example (<https://www.wsj.com/articles/facebook-ad-prices-surge-due-to-barrage-by-democratic-hopefuls-11566984601>), in 2019 during the U.S. Democratic Primary Elections, the cost of reaching likely-Democratic voters and donors on Facebook increased dramatically as the 20 candidates competing for the Democratic presidential nomination drove up demand, with implications for down-ballot candidates trying to reach voters as well. The cost for Republican candidates and organizations to reach voters were significantly less given that there was no competitive Republican presidential primary to drive up demand.

Despite the complexity of advertising price determinations on social media, multiple countries have attempted to regulate in this area:

- Paraguay stipulates that social media platforms that alter their advertising rates in ways that favor any party or political movement over another will be subject to a fine.⁴⁸
- El Salvador’s Electoral Code references a constitutional obligation that the media must provide information on the rates they charge for their services, and that the constitutional principle of equity in pricing among political parties is applicable in the case of social media.⁴⁹
- Venezuelan regulations bar social media platforms from endorsing or supporting candidates while enjoining them from refusing to accept paid advertising from any candidates.⁵⁰

Requiring social media platforms to institute a standard of equity among parties and candidates would require changes to how advertisements are selected and shown to users or how they are priced. Requiring social media platforms to treat candidates and parties equitably presents a range of questions for enforcement, but it is an important principle to consider given platforms' immense power in this regard. Companies have the technological edge to advantage or disadvantage preferred candidates by, for example, more effectively targeting some ads of candidates who have more favorable positions towards the platforms themselves. Recent examples in [India \(https://time.com/5904162/ankhi-das-facebook-india/\)](https://time.com/5904162/ankhi-das-facebook-india/) and the [United States \(https://www.theverge.com/2020/11/1/21544501/facebook-rules-protect-conservatives-instagram-bias-discipline\)](https://www.theverge.com/2020/11/1/21544501/facebook-rules-protect-conservatives-instagram-bias-discipline) have demonstrated the ways in which political pressure and public perception can shape content moderation decisions. Platform actions in this regard would be largely undetectable with the transparency tools available in many countries, and it is uncertain whether such practices would constitute a violation under current legal and regulatory frameworks.

Another possibility is to require that social media platforms publish advertising rates. This type of provision could be incorporated into the standards required of a political ad library or another ad repository, which would allow transparency into the comparative rates that parties and candidates are paying to get their messages out. A movement to create equity in political advertising would likely require increased global pressure from multiple countries – including large markets such as the EU and the U.S. to gain traction, but it is an underexplored avenue. There would also likely be a discussion about how equity should be conceived in light of the different nature of online advertising.

B. COMPEL PLATFORMS TO PROVIDE FREE ADVERTISING SPACE TO CANDIDATES AND PARTIES

The laws and regulations of some countries stipulate that traditional media providers give, in equal measure, [free advertising time \(https://www.lexology.com/library/detail.aspx?g=47546d54-413a-42a1-bdfc-4be3282f041f\)](https://www.lexology.com/library/detail.aspx?g=47546d54-413a-42a1-bdfc-4be3282f041f) or space to political parties or candidates that meet predetermined criteria. This is intended to provide competing parties more equitable access to bring their platforms and ideas to the electorate regardless of their financial resources.

The present study has not identified any jurisdictions that require social media platforms to grant equal free advertising space to candidates or political parties. However, the Bulgarian framework *allows* social media platforms to equitably allocate free advertising space to electoral contestants and requires the platforms to disclose how they allocate it among candidates and parties.⁵¹ The Bulgarian approach could serve as a pilot precursor for countries that contemplate *compelling* social media platforms to offer free campaign advertising space on an equal basis. It is feasible that a national-level provision that draws on existing national law to extend the precedent of equitable free advertising would be able to prevail on major social media companies to provide ad credits to qualified parties, though this is as of yet untested.

LEGAL AND REGULATORY RESPONSES TO DISINFORMATION

5. MEASURES TO PROMOTE DEMOCRATIC INFORMATION DURING CAMPAIGNING AND ELECTIONS (/TOPICS/LEGAL/5-MEASURES- PROMOTE-DEMOCRATIC-INFORMATION- DURING-CAMPAIGNING-AND-ELECTIONS)

Measures to **promote democratic information** are less prevalent, but they do present an opportunity to obligate platforms, and possibly, domestic actors to proactively disseminate unbiased information in ways that can build resilience to political and electoral disinformation. Though there are few real-world examples, this category provides an opportunity to consider what types of legal and regulatory approaches might be feasible.

“Solutions could be aimed at enhancing individual access to information rather than merely protecting against public harm.” — David Kaye (<https://www.the-american-interest.com/2019/06/10/how-not-to-regulate-the-internet/>), United Nations Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression

I. PROMOTE DEMOCRATIC INFORMATION: MEASURES DIRECTED AT DOMESTIC ACTORS

A. REQUIRE PARTIES AND CANDIDATES TO ISSUE CORRECTIONS WHEN PARTY MEMBERS OR SUPPORTERS SHARE BAD INFORMATION

South Africa’s draft code of conduct on *Measures to Address Disinformation Intended to Cause Harm During the Election Period* (detailed discussion of this code of conduct (CoC) can be found in the [topical section on EMB approaches to countering disinformation](https://counteringdisinformation.org/topics/embs/3-emb-codes-conduct-or-declarations-principle-electoral-period) (<https://counteringdisinformation.org/topics/embs/3-emb-codes-conduct-or-declarations-principle-electoral-period>)) stipulates that the Election Commission can compel parties and candidates to correct electoral disinformation that is shared by parties, candidates, or their

members and supporters; “the registered party or candidate shall act immediately to take all reasonable measures in an effort to correct the disinformation and remedy any public harm caused, as may be appropriate in the circumstances and in consultation with the Commission.”⁵²

South Africa’s CoC defines electoral disinformation with specificity and provides a framework for reporting and ruling on violations, which makes these provisions implementable. Definitional specificity around what types of electoral disinformation would be subject to correction and an independent oversight body are necessary for this approach to have an impact and not place undue obligations on political contestants.

If narrowly tailored and enforced, a mechanism to compel political contestants to correct information damaging to the credibility of the electoral process through their own networks of supporters has the potential to reach impacted audiences via the same channels where they might have encountered the problematic content. This, in turn, can amplify messages that election authorities are attempting to disseminate widely.

II. PROMOTE DEMOCRATIC INFORMATION: MEASURES DIRECTED AT PLATFORMS

A. REQUIRE PLATFORMS TO OFFER ELECTION AUTHORITIES FREE ADVERTISING SPACE FOR VOTER EDUCATION

While requiring public or private media outlets to provide equal, free advertising space to political contestants has precedent in a number of countries, another route is to mandate that free advertising space be made available to election authorities. Social media platforms offering free ad-space to election management bodies could be a useful and enforceable provision that could, for example, help boost turnout, educate voters in ways that mitigate invalid voting, or enhance marginalized groups’ access to information.

Using public media channels for this purpose is common practice. In addition, some countries require *private* media actors to offer free space to election authorities. In Mexico, the Constitution stipulates that during electoral periods, radio and television broadcasters must provide 48 minutes of free advertising space every day to be divided between electoral authorities, with space also reserved for messages from political parties.⁵³ Air time is also provided in a more limited amount during non-electoral periods.⁵⁴ Outside of this allotted time, political parties and candidates are not allowed to buy or place any additional television or radio advertisements.⁵⁵ Venezuelan electoral law also requires private television providers to offer free advertising space to the election management body for civic education and voter information.⁵⁶

Obligating private companies to serve as a channel for voter education is an interesting idea. Major platforms, including Facebook, Google, Instagram, and Twitter, have elected to provide voter information, such as election day reminders and instructions on how to vote, of their own volition ([see the subcategory on EMB coordination with social media and technology companies](https://counteringdisinformation.org/topics/embs/7-emb-coordination-technology-and-social-) (<https://counteringdisinformation.org/topics/embs/7-emb-coordination-technology-and-social->

media-companies)). Ahead of the 2020 U.S. elections, Facebook also voluntarily launched a new tool called Voting Alerts (<https://about.fb.com/news/2020/08/launching-voting-information-center/>) that allowed state and local election authorities to reach their constituents with notifications on Facebook, whether or not the Facebook user followed the election authority's Facebook Page. Given the voluntary nature of such measures, voter information integration into the platforms does not take place in all countries or for all elections. Platforms are less likely to roll out features for local or municipal elections, even if those elections are taking place nation-wide, than they are for presidential or parliamentary elections. Considering a requirement for platforms to provide free space for voter education as a part of the legal and regulatory code, particularly in countries where an analogous precedent exists for traditional or public media, could be something to explore. Additionally, requirements for advertisement-funded streaming internet television providers, search engines, or other media intermediaries could also be considered as another place to require advertisements to be integrated.

LEGAL AND REGULATORY RESPONSES TO DISINFORMATION

6. ENFORCEMENT (/TOPICS/LEGAL/6- ENFORCEMENT)

Thoughtful regulation means little if it is not accompanied by meaningful consideration of how that regulation will be enforced. A lack of realism about enforcement threatens to undercut the authority of the regulatory bodies creating provisions and establishes unrealistic precedents for what will be achievable through regulation alone.

The levers of enforcement will change depending on whether provisions are aimed at domestic actors or platforms. In the case of the former, governments and political actors that are in office are increasingly complicit in or actively at fault for participation in the very behaviors that the regulatory actions outlined in this document seek to curb. In these instances, the ability to meaningfully enforce provisions will rely on the independence of enforcement bodies (https://www.ifes.org/sites/default/files/ifes_autonomy_and_accountability_framework_september_2019.pdf) from the executive.

The ability for an individual country to enforce provisions directed at foreign actors is very limited, which is one of the reasons why legal and regulatory approaches directed at foreign actors are not included in this section of the guidebook.

Provisions directed at platforms will vary significantly in how enforceable they may be. Provisions that require alterations to the platform's engineering or global business practices are highly unlikely to come from national-level laws passed in anything other than the largest-market

countries in the world. However, many major social media platforms have thus far been ahead of lawmakers in instituting new provisions and policies to define and restrict problematic content and behaviors or to promote transparency, equity, and/or democratic information. These provisions have not been rolled out equally though, and where national-level legislation might have an impact is in pushing companies to extend their existing transparency tools to the country in question. Platforms will undoubtedly balance their business interests and the difficulty of implementing a measure against the cost of non-compliance with legal provisions in countries where they operate but do not have a legal presence. Recognizing that many countries in the world have limited ability to enforce legal obligations placed on the platforms, legal and regulatory provisions might instead serve to make a country a higher priority for companies as they globalize their ad transparency policies or promote voter information via their products.

6.1 ESTABLISHING WHICH STATE ENTITIES HAVE AN ENFORCEMENT MANDATE

Different institutions may have the right of oversight and enforcement over laws governing the intersection of social media and campaigning, and – given that provisions pertinent to this discussion might be scattered across a legal framework in several different laws – oversight may sit with multiple bodies or institutions. A few common types of enforcement bodies are noted below.

In many countries, responsibility for oversight and enforcement may sit with an **independent oversight body or bodies**. This might be an anti-corruption agency, a political finance oversight body, or a media oversight body, for example. As Germany expands their legal and regulatory framework around social media and elections, implementation and enforcement fall to an independent, non-governmental state media authority. This effort expands the mandate of the body, which has pre-existing expertise in media law, including advertising standards, media pluralism, and accessibility. Analysts of this move to expand German media authorities' scope of work contend that “it is crucial to carefully consider what, if any, provisions could or should be translated to another European context... While Germany's media regulators enjoy a high level of independence (https://cadmus.eui.eu/bitstream/handle/1814/61141/2018_Germany_EN.pdf?sequence=1&isAllowed=y), the same cannot be said of other member states,” citing research that says more than “half of EU member states lack safeguards for political independence in appointment procedures.”⁵⁷

Responsibility for oversight will often be spread across multiple independent bodies or agencies, necessitating coordination and the development of joint approaches. A Digital Regulation Cooperation Forum (<https://www.huntonprivacyblog.com/2020/07/02/ico-teams-up-with-cma-and-ofcom-in-digital-regulation-cooperation-forum/>) has been created in the United Kingdom, for

example, which promotes the development of coordinated regulatory efforts in the digital landscape among the UK Information Commissioner's Office, the Competition and Markets Authority, and the Office of Communications.

Other countries vest **election authorities or election oversight bodies** with the implementation and enforcement capacity of some kind. For election authorities that have political finance, campaign finance, or media oversight mandates, the responsibility to oversee provisions related to social media in elections might, in some instances, be naturally added to these existing capacities. Election authorities may be in the position of having a legal mandate to monitor for violations, or they may have adopted this responsibility independently while lacking authority to enforce. In these instances, legal and regulatory frameworks will need to take into account relevant referral mechanisms to ensure detected violations can be shared with the appropriate body for further action.

In other instances, enforcement sits more directly with the **judicial system**. In the case of France, judges play a direct role (<https://www.gouvernement.fr/en/against-information-manipulation>) in determining what content constitutes information manipulation. In addition to ordering the removal of the manifest, widely disseminated, and damaging content, judges may also order (https://www.loc.gov/law/help/fake-news/france.php#_ftn10) "any proportional and necessary measure" to stop the "deliberate, artificial or automatic and massive" dissemination of misleading information online. In Argentina, the electoral court is responsible for enforcing violations resulting from advertising that takes place outside of the designated campaign period.⁵⁸ Any model that relies on the judiciary to determine what constitutes a violation necessitates a fully independent judiciary with the capacity to understand the nuances of information manipulation and to review and respond to cases quickly.⁵⁹

6.2 BUILDING CAPACITY TO MONITOR FOR VIOLATIONS

Without establishing a capacity to monitor, audit, or otherwise effectively provide oversight, laws, and regulation governing the use of social media during elections are unenforceable. The subsection on *Social Media Monitoring for Legal and Regulatory Compliance* (</topics/embs/4-social-media-monitoring-legal-and-regulatory-compliance>) in the guidebook section on Election Management Body Approaches to Countering Disinformation outlines key questions and challenges in defining a monitoring approach. These include:

- Does the body in question have a legal right to monitor social media?
- What is the goal of the monitoring effort?
- What is the time period for social media monitoring?
- Will the monitoring be an internal operation or conducted in partnership with another entity?
- Does the body in question have sufficient human and financial resources to carry out the desired monitoring effort?

- What social media advertising transparency tools are available in the country?

6.3 CONSIDERATIONS FOR EVIDENCE AND DISCOVERY

The nature of social media and digital content raises new questions in the consideration of evidence and the discovery process. For example, when platforms notify national authorities or make public announcements that they have detected malicious actions on their platforms, it is often accompanied by action to remove the accounts and content in question. When this material is removed from the platform, it is no longer available to authorities that might currently or in the future be capturing the content as evidence of violations of national law.

At present, there does not appear to be a comprehensive obligation on major platforms to preserve and provide information or evidence in the case of an investigation into the origins or financing of content and actions that may be violations of local laws. While in instances of violent crimes, human trafficking, and other criminal acts, major U.S.-based platforms have a fairly consistent record of complying with legal requests by governments for pertinent data, the same does not seem to be true in the case of political finance or campaign violations. A means and precedent for making legally-binding requests for user data from the platforms when a candidate or party is under credible suspicion of violating the law is an essential route to explore for enforcement.

Granted, the platforms also play a critical role in ensuring user data gathered on their platforms is not handed over to government actors for illegitimate purposes. The determination of what does and does not constitute a legitimate purpose is one that necessitates careful deliberation and the establishment of sound principles. There is also likely to be frequent conflict between what platforms deem to be requests for data with the potential for abuse and what the national authorities requesting that data might think. Particularly



HIGHLIGHT

In instances where a case is being brought against an actor for illegal conduct on social media, a legal request to preserve posts and data may be a step that authorities or plaintiffs need to consider. Dominion Voting Systems, for example, has pursued this action in a series of defamation cases against media outlets and others for falsely claiming that the company's voting machines were used to rig the 2020 U.S. elections. Dominion sent letters to Facebook, YouTube, Parler, and Twitter requesting that the companies preserve posts (<https://www.washingtonpost.com/technology/2020/11/12/dominion-voting-social-media-letters/>) relevant to their ongoing legal action.

for countries that have leaned heavily into the use of their criminal code to sanction problematic speech, the platforms may preserve legitimate resistance to complying with requests for user data that have a high potential for abuse.

6.4 AVAILABLE SANCTIONS AND REMEDIES

Countries have used a variety of sanctions and remedies to enforce their legal and regulatory mandates. Most of these sanctions have precedent in existing law as it pertains to analogous offline violations.

The issuing of **fin**es for political finance or campaign violations has a well-established precedent. In the context of violations of digital campaigning rules, fines are also a common sanction. Argentinian law, for example, stipulates that fines will be issued to human or legal entities that do not comply with content and publication limits on advertisements, including those transmitted via the internet. Argentina's law assesses the fine in relation to the cost of advertising time, space, or internet bandwidth at the time of the violation.⁶⁰

Fines can also be directed at social media companies or digital service providers that do not meet their obligations. Paraguay, for example, holds social media companies vicariously liable and subject to fines for breach of campaign silence, illicit publication of opinion polls, or for engaging in biased pricing.⁶¹ It is unclear if Paraguay has successfully levied these fines against any social media companies.

Some legal and regulatory frameworks carry the threat of **revoking public funding** as a means of enforcement. In contrast to the penalty of a fine for individuals in breach of the law, the Argentinian Electoral Code stipulates that *political parties* that do not comply with limitations placed on political advertising will lose the right to receive contributions, subsidies, and public financing for a period of one to four years.⁶² The effectiveness of this sanction is heavily dependent on the extent to which parties rely on public funding for their income.

Provisions might seek to remedy harm by requiring entities found to be in violation of the law to **issue corrections**. As referenced in the section on promoting democratic information, South African regulation stipulates that the election commission can compel parties and candidates to correct electoral disinformation that is shared by parties, candidates, or their members and supporters. However, mandates to provide corrections can be manipulated to serve partisan interests; Singapore's Protection from Online Falsehoods and Manipulation Act in 2019, which has been subject to heavy criticism for its use to silence opposition voices, requires internet service providers, social media platforms, search engines, and video-sharing services like YouTube to issue corrections or remove content if the government deems it false and that a correction or removal is in the public interest. The law specifies that a person who has communicated a false

statement of fact may be required to make a correction or remove it even if the person has no reason to believe the statement is false.⁶³ Individuals who do not comply are subject to fines up to \$20,000 and imprisonment.⁶⁴

Another sanction is the **banning of a political party or candidate from competing in an election**. The Central Election Commission of Bosnia and Herzegovina fined and banned a party (https://balkaninsight.com/2020/10/08/serbian-party-banned-from-bosnian-election-over-hateful-video/?utm_source=Balkan+Insight+Newsletters&utm_campaign=0ad9aea5ec-BI_DAILY&utm_medium=email&utm_term=0_4027db42dc-0ad9aea5ec-319799297) from participating in 2020 elections for sharing a video that violated a provision against provoking or inciting violence or hatred,⁶⁵ though this decision was overturned (<https://www.sarajevotimes.com/170077-2/>) by the courts upon appeal. This sanction is at high risk of political manipulation and, if considered, must be accompanied by sufficient due process and a right of appeal.

In some instances, enforcement has resulted in the **annulment of election results**. The Constitutional Court of Moldova annulled a mayoral election (<https://www.rferl.org/a/moldovans-protest-nullification-chisinau-mayoral-election/29316498.html>) in the city of Chisinau because both competitors were campaigning on social media during the campaign silence period. In the aftermath of this decision, which was viewed by many as disproportionate to the offense, Moldovan regulators introduced a new provision allowing campaign materials on the internet which were placed before Election Day to remain visible. Election annulment is an extreme remedy that is highly vulnerable to political manipulation and should be considered in the context of international best practice on validating or annulling an election (<https://www.ifes.org/publications/when-are-elections-good-enough>).

Countries have **banned or threatened to ban access to a social media platform** within their jurisdiction as a means to compel compliance or force concessions from global social media platforms. The Government of India, for example, threatened to ban WhatsApp (<https://www.indiatimes.com/technology/news/indian-government-will-ban-whatsapp-in-country-if-the-app-doesn-t-find-a-way-to-trace-hoaxes-353391.html>) in 2018 following a string of lynchings resulting from viral rumors being spread via the messaging application. WhatsApp refused to accede to the government's demands on key privacy provisions but did make alterations to the ways in which messages were labeled and forwarded within the app in response to government concerns. India also banned TikTok (<https://www.nytimes.com/2020/06/29/world/asia/tik-tok-banned-india-china.html>), WeChat, and a range of other Chinese apps in 2020. In 2018, the Indonesian government banned TikTok (<https://www.reuters.com/article/us-indonesia-bytedance/indonesia-overturns-ban-on-chinese-video-app-tik-tok-idUSKBN1K10A0>) for several days on the basis that it was being used to share inappropriate content and blasphemy. In response, TikTok quickly acceded to the government's demands and began censoring such content. The Trump administration threatened to ban TikTok in the United States over data privacy concerns unless the Chinese-owned company sold its U.S. operations. In 2017, Ukrainian

President Petro Poroshenko signed a decree that blocked access (<https://www.theguardian.com/world/2017/may/16/ukraine-blocks-popular-russian-websites-kremlin-role-war>) to a number of Russian social media platforms on national security grounds.

Banning access to entire platforms as a means to force concessions from companies is a blunt-force approach that is only likely to yield results for countries with massive markets of users. Far more frequently, bans on social media platforms have been used as a tool by authoritarian leaders to restrict access to information among their populations.

Regulating social media in campaigning, particularly in a way intended to deter or mitigate the impact of disinformation, is far from coalescing around established and universally accepted good practices. As countries take legal and regulatory steps to address disinformation in the name of protecting democracy, the uncertainty and definitional vagueness of key concepts in this space has the potential to result in downstream implications for political and civil rights. Concerns about free speech, for example, are elevated when content is removed without any judicial review or appeals process. Critics point to the dangers of allowing unaccountable private social media companies and digital platforms to decide what content does or does not comply with the law. For example, if sanctions are severe, it might incentivize companies to overcorrect by removing permissible content and legitimate speech. The existence of **robust appeals mechanisms** is essential for preserving rights.

EXPOSING DISINFORMATION THROUGH ELECTION MONITORING

0. OVERVIEW - ELECTION MONITORING (/TOPICS/MONITORING/0-OVERVIEW- ELECTION-MONITORING)

Written by Julia Brothers, Senior Advisor for Elections and Political Processes at the National Democratic Institute

Democratic elections rely on a competitive process, faith in electoral institutions, and informed participation by all citizens. However, the deployment of false, exaggerated, or contradictory information in the electoral environment has been effective in undermining these principles around the world. By interfering with the formation and holding of opinions, disinformation amplifies voter confusion, dampens turnout, galvanizes social cleavages, advantages or disadvantages certain parties and candidates, and degrades trust in democratic institutions. While anti-democratic disinformation campaigns are not new, modern information technology and the

platforms by which citizens get their news, including online and via social media, encourage information dissemination at speeds, distances, and volumes unprecedented in preceding electoral cycles.

International standards for democratic elections assure open, robust, and pluralistic information environments that promote equal and full participation in elections by citizens and contestants alike. These standards are enshrined in international and regional instruments, which reflect pre-existing, globally-recognized commitments that pertain to disinformation, including:

- **The rights to hold opinions and to seek and receive information in order to make an informed choice on election day:** Everyone has the right to form, hold, and change opinions without interference, which is integral to freely exercising the right to vote.¹ Voters also have the right to seek, receive, and impart accurate information that allows them to make informed choices regarding their future, free from intimidation, violence, or manipulation.² Further, institutions are generally obligated to be transparent regarding electoral information so that voters can be informed and data sources can be held accountable.³ These rights are enshrined for all citizens regardless of race, gender, language, area of origin, political or other opinion, religion, or other status.⁴ Increasingly, organizations are working to link these standards with principles focused on disinformation and cyberspace (/node/2743/). Electoral related disinformation efforts subvert these rights, because they are designed to overwhelm genuine political debate by intentionally deceiving voters, creating confusion, exacerbating polarization, and undermining public confidence in the electoral process.
- **The right to a level playing field:** Universal and equal suffrage, in addition to voting rights, include the right to seek to be elected to public office without discrimination. Governments' obligations to ensure level playing fields for electoral contestants are derived from this norm. The UN Human Rights Committee provides guidance on this in its General Comment 25 (<https://www.equalrightstrust.org/ertdocumentbank/general%20comment%2025.pdf>) to the ICCPR. The norm implies providing security from defamatory attacks and other forms of false information aimed to harm a candidate's or a party's electoral fortunes. The obligations extend to government-controlled media, and the norm applies to professional ethics for journalists and private media.⁵ Fact-checking, other forms of verification, and traditional and social media monitoring relate to this norm, as well as to voters' rights to receive accurate information upon which to make informed electoral choices. Manipulation of the information environment can undermine equitable competition, particularly for those that are disproportionately impacted by disinformation campaigns, like women and marginalized communities, who already face an uneven playing field.
- **Freedom of expression, the press, and regulation:** The aforementioned commitments must be balanced by the freedoms of everyone to hold opinions and to express them, including the need to respect and protect a free press. One aspect of addressing disinformation campaigns is to develop proper legal and regulatory frameworks, including effective sanctions. Gendered, racial, ethnic, religious, and other forms of hate speech and incitement to violence are often diffused throughout disinformation campaigns affecting candidates and voters alike. Legal regulations in this area, like protection of personal

reputation, can be applicable in the disinformation context. However, regulation should not be overemphasized, and care is needed to safeguard freedom of expression while trying to protect the integrity of the information space in elections and beyond them. The UN Human Rights Committee provides guidance on this in General Comment 34 (<https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf>).

Recognizing these necessary democratic conditions, the existence and impact of disinformation must be considered in any comprehensive assessment of an electoral process. Even if an election is well-organized and transparent, a highly compromised information environment leading up to and on election day can subvert its credibility. Identifying the types, volumes, and patterns of mis- and disinformation that may affect electoral integrity is crucial for mitigating their impact. Political watchdogs should analyze deficiencies in the information environment with an understanding of the social norms and cleavages in the local context when determining the integrity of an election and creating accountability for the universe of stakeholders who engage in or benefit from disinformation tactics.

Traditional electoral safeguards, particularly election observers, are expanding their capacities, activities, relationships, and advocacy efforts to confront disinformation threats to electoral integrity. Debunking fake news through emerging networks of fact checkers and bolstering media and digital literacy play important roles in building resilience and enhancing the information environment around elections. Those actions, as well as robust efforts to properly inform political debate and provide accurate electoral data, can inoculate against information disorder. All of such efforts can complement each other to safeguard electoral and political processes.

EXPOSING DISINFORMATION THROUGH ELECTION MONITORING

1. RESPONDING TO THE DISINFORMATION THREAT THROUGH ELECTION MONITORING: PROGRAMMING APPROACHES AND CATEGORIES (/TOPICS/MONITORING/1-RESPONDING- DISINFORMATION-THREAT-THROUGH- ELECTION-MONITORING-PROGRAMMING)

Election monitoring programs broadly serve to promote electoral integrity through enhanced participation, inclusion, transparency, and accountability, thus fostering citizen empowerment and confidence in the democratic process.

Developing the right election observation intervention(s) to respond to disinformation should not be done without first considering the context of each electoral environment.

Decisions to use technologies and methodologies should be made through an inclusive process, with consideration of the accessibility and technology gaps among different groups of observers and citizens, including along gender, age, geography, and other lines. In addition, [identifying and exposing online barriers for women \(/node/13/\)](#) and marginalized groups in electoral processes necessarily requires an inclusive, gender sensitive approach and may require observers to incorporate and balance specialized methodologies into their overall effort that create an accurate picture of how the electoral landscape affects specific populations.

There are several options to address the specific threats that disinformation poses to electoral integrity in an individual country context:

- **Citizen election observation to identify and expose disinformation as it relates to electoral integrity**, including monitoring online and traditional media around an electoral process
- **International election observation of the electoral information environment, including disinformation, in the short and long-term** by credible international and regional observation missions and in line with the Declaration of Principles for International Election Observation
- **Advocacy for norms, standards, and policies to address disinformation in elections**, including efforts by civil society and/or other groups to advocate for a range of appropriate responses from social media platforms and other private sector actors, legal reforms, policies and resource allocation from governments or legislatures, and support for norms-building and standards ([/node/2743/](#)) from regional and international instruments to combat disinformation during elections.
- **Building more effective partnerships between election observers and other key stakeholders**, such as civic tech groups, fact-checkers, journalists, media monitors, electoral management bodies, women's rights organizations and other CSOs that are composed of and represent marginalized groups, etc.
- **Knowledge-sharing and developing best practices around combating disinformation in elections** through workshops, online exchanges, guidance notes and other information



DESIGN TIP

The nature, vulnerabilities, mitigating factors, and opportunities around the electoral information, online and otherwise, vary significantly from country to country, and successful projects have demonstrated the importance of conducting a preliminary assessment to identify these factors before designing a program. Subsequently, **monitoring methodologies and approaches should be shaped and driven by objectives and organizational capacity, not by available tools.**

sharing forms.

These interventions are explored in more detail below and demonstrate how focused electoral observation and analysis can enhance accountability and neutralize disinformation threats. Election monitoring is ideally conducted throughout the pre-election, election day, and post-election periods to evaluate all relevant aspects of the electoral process. Many of the case studies highlighted in this chapter are not standalone projects, but are part of broader election monitoring efforts that include online monitoring as a distinct component.

EXPOSING DISINFORMATION THROUGH ELECTION MONITORING

2. CITIZEN ELECTION OBSERVATION: MONITORING ONLINE AND OFFLINE CONTENT IN THE ELECTORAL CONTEXT (/TOPICS/MONITORING/2-CITIZEN-ELECTION-OBSERVATION-MONITORING-ONLINE-AND-OFFLINE-CONTENT-ELECTORAL)

Election observers frequently adjust their methodologies to meet evolving tactics that undercut credible electoral processes, often in the pre-election period. Citizen election monitors, who are often viewed as trusted, politically impartial voices, are well-equipped to investigate, expose, and mitigate the effects of information manipulation around elections. They understand online vernacular and the significance of slang and other terms that are key to identifying disinformation and its connections to hate speech, incitement, and other means of fanning social divisions. That understanding can be helpful to international election observers and foreign researchers. Moreover, national organizations can provide ongoing monitoring not only during elections, but also during major legislative votes, national plebiscites, and the period between elections when the online manipulation of political narratives tends to take root.

For example, in Georgia, the citizen election observer group [International Society for Fair Elections And Democracy \(ISFED\)](http://www.isfed.ge/eng) (<http://www.isfed.ge/eng>) developed a multi-prong approach to identify disinformation tactics designed to influence voters and subvert fact-based discourse ahead of the 2018 presidential election and subsequent run-off.



HIGHLIGHT

Using an NDI-designed tool (the Fact-a-lyzer) that was created specifically for citizen observers to monitor platforms like Facebook and Twitter, ISFED monitored a range of electoral integrity issues on social media, including abuse of state resources, campaign finance, the strategic spread of disinformation and divisive narratives, and the use of official and unofficial campaign pages in the elections to discredit candidates and, in some cases, civil society organizations. Some of their findings, including clear campaign finance violations, were flagged for government oversight institutions that subsequently levied fines on the violators. In addition, through social media monitoring ISFED was able to identify a number of suspicious fake media pages that Facebook eventually removed in a high-profile operation for coordinated and inauthentic behavior. The group continued to monitor social media between the 2018 presidential elections and the 2020 parliamentary polls, identifying a series of Kremlin-backed disinformation campaigns. (<https://isfed.ge/eng/sotsialuri-mediis-monitoringi>)

While fact-checking groups and other media integrity initiatives serve critical functions in weeding out false and misleading narratives, social media monitoring by citizen election observers tends to have different goals and timelines. The objective is not to quickly verify and/or invalidate individual stories but rather to identify and evaluate the impact information trends may have on electoral integrity, build accountability around a variety of actors participating in the electoral process, and provide actionable recommendations.

Problematic pages highlighted by ISFED were all in Georgian, a language not widely spoken outside of the country and even less common among tech platform content moderators. The prevalence of disinformation in local language content reinforced the importance of citizen monitoring to appreciate linguistic subtext and more easily interpret social media content and behavior within the electoral context. ISFED's effort has been rooted in long-term monitoring with well-trained staff and access to advanced data collection tools like Fact-a-lyzer and Facebook's Crowdtangle, which have improved their capacity and ability to perform more advanced research. Their social media monitoring effort is ongoing (<https://isfed.ge/eng/sotsialuri-mediis-monitoringi>) to capture inter-election trends and identify how some narratives developed online well in advance of an election become weaponized for electoral advantages or disadvantages. Such an ambitious approach requires long-term resources and access to bulk public content.

In Nigeria, election monitors broadened traditional fact-checking efforts to conduct more nuanced research to identify underlying information trends ahead of the 2019 General Elections. NDI partnered with the Centre for Democracy and Development — West Africa (CDD-West Africa) (<https://www.cddwestafrica.org/>), which was already undertaking a robust media literacy and fact-checking campaign to quantitatively analyze the information environment in the weeks leading up to the elections. NDI hired Graphika, a private research firm that conducts data collection and analysis on online platforms such as Facebook and Twitter, to provide much of the research support. Through the combination of Graphika's analysis and the manual data collection of the fact-checkers, CDD-West Africa was able to highlight the depth and scope (<https://www.cddwestafrica.org/wp-content/uploads/2019/07/SORTING-FACT-FROM-FICTION.pdf>)

of certain narratives around the elections, particularly related to Islamophobia and foreign influence. It also uncovered coordinated fake news networks and signs of inauthentic automated accounts.

These efforts were complemented by research (<https://www.cddwestafrica.org/wp-content/uploads/2019/07/WHATSAPP-NIGERIA-ELECTION-2019.pdf>) CDD-West Africa conducted in partnership with the University of Birmingham examining the use of WhatsApp ahead of the elections. CDD-West Africa briefed a number of international election observation missions on their findings, which contributed to election day statements and further analysis. By augmenting their fact-checking efforts with sophisticated data analysis, CDD-West Africa was able to spot broad trends impacting the electoral process while still providing updates on the online environment in real time.

Penplusbytes (<http://penplusbytes.org/>), a local NGO in Ghana, developed a Social Media Tracking Center (<http://africanelections.org/ghsmtc/>) (SMTC) for the 2012 Ghanaian presidential elections and revived it for the 2016 presidential elections to identify electoral malpractices as they occur, using such information to warn relevant institutions and stakeholders quickly. The Penplusbytes teams used the Aggie social media tracking software (<https://www.getaggie.org/>) developed by the Georgia Institute of Technology and the United Nations University to monitor and verify instances of misinformation on Facebook and Twitter. They passed relevant information on to the National Elections Security Task Force (NESTF) which took action based on their findings.

In Colombia, the civic group Electoral Observation Mission (Misión de Observación Electoral or MOE) (<https://moe.org.co/mision/#1488909074333-b43b2e85-74f4>) has been monitoring online aspects of electoral processes since the referendum on the country's peace agreement held in 2016. In many ways, the peace process has helped define Colombian society in recent years, as it fights to consolidate its progress democratically, reconcile various combatants in the war, integrate rebels back into society, and ultimately avoid regression into the conflict that ravaged the country for decades. According to MOE's Director of Communications, Fabian Hernandez: "Just at that moment, MOE made the first analysis of social media. Our focus at that time was to look at how much electoral crimes were talked about online, what were the arguments with which people spoke of a referendum, [and was it to be] an endorsement of peace? We did not foresee, we did not envision that misinformation was going to be such a serious problem. Therefore it was not the object of our study, but we had a tool to give us alerts and others...that the great risk to the referendum was misinformation, how it was circulating through WhatsApp and text messages, and through Instagram, but also Twitter, a lot of false information, misinformation or exaggerated or decontextualized information that ended up being false."⁶

Subsequently, MOE developed more sophisticated, data-driven social media research plans, linkages to platforms for reporting, and other advanced forms of coordination. During the 2018 presidential election, MOE worked to develop online data collection methods and mechanisms for reporting to the platforms and electoral authorities. As Hernandez noted: "After Brexit, Colombia was a very interesting pilot for the world of how disinformation could change elections. And that made our approach for the study of social media by 2018 characterizing disinformation. That is why we came to the study of who produces misinformation and how misinformation becomes

viral.”⁷ With the help of social listening platforms, MOE collected data around keywords from Facebook, Instagram, Twitter, Youtube, blogs and other media, recording nearly 45 million pieces of content (<https://moe.org.co/wp-content/uploads/2019/03/2.-Monitoreo-de-Redes-Sociales-Intolerancia-y-Noticias-Falsas.pdf>). This content was analyzed with natural language processing software to contribute to a final report covering both rounds of the election, as well as congressional and intra-party discussion rounds.

Local elections are similarly vulnerable to misinformation and disinformation campaigns, but often receive less scrutiny and attention from international actors, the media, and researchers, further elevating the importance of citizen watchdog organizations. As noted by Hernandez: "In local elections we had the same exercise of looking at social media, and today our analysis focuses on: Disinformation, hate speech, intolerance or aggressiveness; and finally xenophobia, immigration, and Venezuela. From the traditional media it was understood that people with less education were more vulnerable to manipulation, which are barriers placed by the type of education, because of the little education they receive, that is why misinformation was easier."⁸

Citizen observation groups are more likely to capture digital threats at the local level than their international counterparts. They have a stronger understanding of what is said and what is meant on social media and insight into the particular experiences of women, members of other marginalized groups, and other populations online at the local, regional and national level. Integration of these perspectives is essential to informing the monitoring process. Moreover, national organizations can provide ongoing monitoring not only during elections, but also during major legislative votes, national plebiscites like Colombia's over the peace process, and the period between elections when the manipulation of online political narratives tends to take root. Linking traditional observers such as MOE with other kinds of online monitoring organizations, digital rights groups, fact checkers, civil society representing women and marginalized groups, and civic technologists becomes critical to understand the complete picture of a country's social media landscape over time.

EXPOSING DISINFORMATION THROUGH ELECTION MONITORING

3. INTERNATIONAL ELECTION OBSERVATION OF THE ELECTORAL INFORMATION ENVIRONMENT (/TOPICS/MONITORING/3-

INTERNATIONAL-ELECTION-OBSERVATION-ELECTORAL-INFORMATION-ENVIRONMENT)

International election observation missions are committed to assessing the quality of an electoral process in its entirety, including in the pre-election, election day, and post-election periods. This commitment is rooted in the Declaration of Principles for International Election Observation (Declaration of Principles or DoP). Therefore, consideration of the information environment, including the role of disinformation, hate speech, and other online forms of content where they play a significant role represent a critical part of any mission assessment. Additionally, according to the DoP, gender considerations must be emphasized not only at the individual mission level but also at the international and normative level. In the context of the information environment, this would include an understanding of the dimensions of Violence Against Women in Politics (VAW-P) and in Elections including their online manifestations such as gendered disinformation (/node/13/). This may involve incorporating analysis and recommendations concerning the information environment into pre-election and election day statements. Missions should strive to expand the pool of key informants and interlocutors from whom long- and short-term observers collect information, such as social media experts, academics, tech industry representatives, women's rights activists, and media monitors, both in-country and from outside. Observation missions may also want to diversify the profiles of pre-election and election day analysts and delegates to include civic technologists, digital communications experts, or others with particular knowledge of gendered digital manipulation techniques. Where needed, missions may seek to influence social media firms if analysis reveals serious challenges to electoral integrity, whether through disinformation, hate speech or other influences.

In some cases, particularly for missions in countries experiencing acute disinformation campaigns around elections, a core team member or analyst could be slotted to concentrate on developing analysis of the dimensions of disinformation in the electoral context. For instance, in Nigeria the European Union (https://eeas.europa.eu/election-observation-missions/eom-nigeria-2019_en) deployed a media and digital communications analyst to cover the online space for the 2019 Nigerian presidential election, and has deployed other media monitors in different contexts globally.

Similarly, for its international election observation missions of Ukraine's 2019 presidential and parliamentary elections, NDI hired a long-term information environment analyst as part of the mission's core team. The mission recognized the role that the information environment, including disinformation in traditional and social media, was likely to play in those high-profile elections. Like the mission's other thematic experts, such as gender and legal framework analysts, the information environment analyst provided a clear focal point on the issue to ensure that all aspects of the mission were taken into account, including information disorder as an electoral integrity issue.

The long-term analyst (LTA) collected data from key interlocutors and pre-existing data sets and monitored 26 regional and national Telegram channels, which revealed a pattern of disproportionately negative posts regarding the electoral process and the two major presidential candidates. This and other analyses by the LTA contributed substantially to the [findings](https://www.ndi.org/sites/default/files/NDI%20Ukraine%20-%20March%2031%202019%20Presidential%20Election%20-%20Election%20Observation%20Statement%20-%20Press%20Conference%20Final%20ENG%20vf2.pdf) (<https://www.ndi.org/sites/default/files/NDI%20Ukraine%20-%20March%2031%202019%20Presidential%20Election%20-%20Election%20Observation%20Statement%20-%20Press%20Conference%20Final%20ENG%20vf2.pdf>) of the [observation](https://www.ndi.org/publications/statement-ndi-election-observation-mission-ukraines-april-21-2019-second-round) (<https://www.ndi.org/publications/statement-ndi-election-observation-mission-ukraines-april-21-2019-second-round>) [mission](https://www.ndi.org/publications/statement-ndi-election-observation-mission-ukraine-s-july-21-2019-snap-parliamentary) (<https://www.ndi.org/publications/statement-ndi-election-observation-mission-ukraine-s-july-21-2019-snap-parliamentary>). In particular they framed the extent to which foreign and domestic online campaigns influenced the electoral process and how political parties, candidates, and less transparent third party supporting accounts utilized online campaigning to shape the digital landscape. This builds on the NDI's experience from its 2017 observation mission in Georgia, during which it deployed a long-term information environment analyst for the first time (<https://cesko.ge/res/docs/NDI-GEEOM2017-IR-ENG.pdf>).

Other international and intergovernmental observer organizations, such as the [Carter Center](https://cyber.harvard.edu/sites/default/files/2019-11/Comparative%20Approaches%20to%20Disinformation%20-%20Michael%20Baldassaro%20Abstract.pdf) (<https://cyber.harvard.edu/sites/default/files/2019-11/Comparative%20Approaches%20to%20Disinformation%20-%20Michael%20Baldassaro%20Abstract.pdf>), [Democracy Reporting International](https://democracy-reporting.org/wp-content/uploads/2019/12/Final-version_Online-Disinformation-Risk-Assessment_Presidential-elections-in-Croatia-2019-2020.pdf) (https://democracy-reporting.org/wp-content/uploads/2019/12/Final-version_Online-Disinformation-Risk-Assessment_Presidential-elections-in-Croatia-2019-2020.pdf), the [Organization of American States \(OAS\)](https://www.oas.org/en/iachr/expression/publications/Guia_Desinformacion_VF%20ENG.pdf) (https://www.oas.org/en/iachr/expression/publications/Guia_Desinformacion_VF%20ENG.pdf), and the [OSCE/Office for Democratic Institutions and Human Rights \(ODIHR\)](https://www.ifes.org) (<https://www.ifes.org>) have also been integrating social media monitoring into their broader observation missions over the last several years, and the international observation community continues to work together to [strengthen capacity and harmonize norms in this area](/node/2743/) (</node/2743/>). In some cases, they collaborate with civil society organizations such as Slovakia's Memo 98, which has developed media monitoring programs since the 90s. The linkage between traditional media monitoring and social media monitoring is important to note, and Memo 98, as with many organizations, has shifted from examining traditional media to social networking platforms in the last five years. Since its initial forays examining the online reach of Russian outlets such as RT and Sputnik in 2015, Memo 98 has broadened its social media focus, supporting the European Union, OSCE, and other election observation missions in Europe and elsewhere.

Memo 98 media monitoring activists deployed online analysis through OSCE monitoring missions in Georgia in 2017 and for the [European Union Parliamentary Elections](http://memo98.sk/article/parties-focused-more-on-domestic-politics-than-eu-topics) (<http://memo98.sk/article/parties-focused-more-on-domestic-politics-than-eu-topics>). In the latter case, they worked to determine the extent to which messages on Facebook impacted the issues presented by political parties during the election. Memo 98 did not find that the parties attacked each other significantly in the posts and, rather, resulted in the unification against extremism. Memo 98 also monitored the Belarus 2020 election in collaboration with Belarusian NGOs Linked

Media and the EAST Center. They developed reporting focused on social media and contrasted how the country's skewed national traditional media resulted in President Lukashenko receiving 97 percent of coverage (<http://memo98.sk/article/election-monitors-lukashenka-used-every-trick-in-his-outdated-playbook-to-survive>) while opposition candidates were able to post and garner some attention on social media such as Facebook (<http://memo98.sk/article/television-and-social-media-monitoring-presidential-election-in-belarus-2020>).

As with other groups in Eastern Europe, Memo 98 is uniquely positioned to understand the potential of foreign influence operations, particularly emanating from Russia. As its director, Rasto Kuzel, notes:

"Obviously we could not ignore [the online space] any more after 2016. And that's why we started working on some kind of methodological approach. We saw that...understanding the basics of content analysis, understanding what the data shows us, understanding the larger picture, you show some of these infinitives but do we get a sense? Like how big a problem this is in the whole election environment. I mean, what is the real impact of social media in a particular country? And how does it correlate with traditional media and so on and so forth."⁹

Balancing the impact of both social and traditional media is a challenge in understanding conversations online, where the traditional media also plays a role. While traditional media monitoring is limited to the officially licensed media, television, radio and print, social media is difficult, if not impossible, to observe completely. Yet, as Kuzel notes, it is important to include it in any observation, and groups are working collectively to develop new methodologies for the online environment. With the Council of Europe, Kuzel has recently published a guide on media monitoring in elections that includes a section on social media methods (<https://rm.coe.int/monitoring-of-media-coverage-of-elections-toolkit-for-civil-society-or/1680a06bc6>) based on his experience. New tools like Crowdtangle, a social media research application owned by Facebook that collects publicly available data about groups and pages on Facebook, Instagram, Twitter and Reddit, form a critical component. As Kuzel notes: "With Crowdtangle for Facebook and Instagram we can get the historical data, which makes a big difference. We feel more comfortable when we can analyze bigger periods and more data and that was not always the case."¹⁰ Tools such as Crowdtangle increase the field of view for observers, but hide comments and other private information about users. Observers and other researchers should be aware of any tool's blind spots, (e.g. private Facebook groups) that are not covered by the platform.

EXPOSING DISINFORMATION

THROUGH ELECTION MONITORING

4. ADVOCACY FOR NORMS, STANDARDS, AND POLICIES TO ADDRESS DISINFORMATION IN ELECTIONS (/TOPICS/MONITORING/4-ADVOCACY-NORMS-STANDARDS-AND-POLICIES-ADDRESS-DISINFORMATION-ELECTIONS)

Electoral monitoring and electoral reform initiatives present a number of opportunities for advocacy at the local, national, and international level. Election observers are in a strong position to provide clear and actionable recommendations through observation statements as well as long-term electoral reform projects to enhance transparency and promote a healthy electoral information environment. Election observation statements by international election observers can draw international attention to particular challenges, and recommendations within those statements often serve as benchmarks for democratic actors to pursue advances and create accountability for their relevant targets. For instance, international election observation missions in Ukraine noted ongoing shortcomings of the tech industry in online political advertising transparency and limitations in their ability to manage electoral disinformation at the local level (https://www.ndi.org/sites/default/files/NDI%20Ukraine%20-%20July%2021%202019%20Parliamentary%20Election%20Observation%20Statment%20-%20ENG%20v_0.pdf). International organizations that observe elections also can draw attention to normative issues to be addressed by technology companies and can help gain the attention of intergovernmental organizations and other sectors concerning those issues.

Meanwhile, citizen election observers already play effective roles in highlighting deficiencies in regulations and enforcement in their own countries and advocating for reforms. Amidst ongoing attempts by political groups and foreign actors to undermine the election environment in Georgia, ISFED coordinated with 48 other leading Georgian civil society and media organizations to successfully pressure Facebook to increase transparency and accountability measures (<https://agenda.ge/en/news/2020/2255>) ahead of the 2020 parliamentary elections. Citizen observer groups Sri Lanka worked together to pressure the government to provide stronger campaign finance oversight mechanisms for political ads online.

Supporting electoral reform efforts and dialogue between election management bodies (EMBs) and observers to expand the availability of election information and encourage transparency of political data, such as voting results (from the polling station to the national level), voter registries and related population numbers, procurement processes, complaints adjudication, and political advertisements on social media, can be central to address misinformation and disinformation. Transparent, accessible data can inoculate EMBs from conspiracy theories or misinformation while increasing citizens' ability to fact-check information they may receive from third parties.

Constructive engagement on this front can help build public confidence in otherwise vulnerable electoral institutions, and encourage EMBs to develop their own strategies for mitigating and responding to disinformation attempts to undermine their own credibility

EXPOSING DISINFORMATION THROUGH ELECTION MONITORING

5. BUILDING BETTER PARTNERSHIPS (/TOPICS/MONITORING/5-BUILDING-BETTER-PARTNERSHIPS)

Disinformation can manifest in complex ways and may require a range of actors to address. Observer groups that lack the time, resources or skills to launch their own social media monitoring efforts may also collaborate, formally or informally, with media monitoring groups, academics, tech advocates, journalists' associations, women's rights organizations, organizations that are comprised of and represent marginalized groups, conflict prevention organizations, or other actors that may already be examining disinformation issues. Such partnerships can ensure that election observers give due consideration of the quality of the electoral information space in their overall electoral analysis without conducting direct data collection themselves.

Observers may also consider partnerships with nontraditional monitoring groups, such as fact checkers and other research organizations with experience in social media and broader online monitoring. A report coauthored by the Open Society European Policy Institute and Democracy Reporting International highlights how groups ranging from academic projects (e.g. the Oxford Internet Institute's Computational Propaganda Project and the Brazilian Getúlio Vargas Foundation's Digital Democracy Room), think tanks (e.g. Atlantic Council's Digital Forensic Research Lab) to fact-checking organizations (e.g. debunk.eu), to the private sector (e.g. Bakamo.Social) have all contributed to election monitoring in various forms.¹¹ Multi-stakeholder collaboration forms one potential basis for development of next generation election observation and monitoring, allowing election observers to incorporate the findings of credible partners into electoral assessments rather than duplicate their work, thus expanding the potential leverage for advocacy around norms and standards. This may be a particularly useful approach for international election observers, who are outsiders by definition and who conduct analysis over a relatively short timeframe.

Relationships with credible EMBs are particularly important for both addressing disinformation through voter education and for encouraging EMBs to enhance their abilities to rapidly respond to electoral disinformation (/node/31/). For example, NDI co-hosted an



DESIGN TIP

event with Mexico's election commission (INE) that focused on responding to disinformation threats in Mexico's July 1, 2018 elections. This brought together a diverse mix of electoral stakeholders, including representatives from major tech platforms, academics, election monitors, and other civic activists in addition to election administrators. To facilitate further collaboration between electoral stakeholders, NDI organized workshops and regular coordination meetings between civic tech groups, fact-checkers, and citizen election observer groups to collaborate on combatting electoral disinformation. This approach was particularly helpful in merging Mexico's civic tech expertise with the electoral analysis lens that observer groups could provide.

Groups can also leverage partnerships to convene multi-stakeholder roundtables about countering disinformation, or to expand the agenda for pre-existing fora for sharing information around elections to also discuss how parties, media, election management bodies (EMBs), observers, and others can help one another spread the information to the broader public and create accountability for maintaining information integrity.

Following a similar model, the Taiwan Foundation for Democracy (TFD), under the Global Cooperation and Training Framework (GCTF) mechanism, organized a conference in September 2019 entitled "Defending Democracy Through Promoting Media Literacy II."¹² Its purpose was to examine the different ways disinformation influences elections around the world, the implementation of media literacy education in curricula, how government and civil society initiatives have evolved to combat disinformation, and the challenges they face. The conference recognized the fact that Taiwan and other countries in the Asia-Pacific region faced presidential and general elections in 2020.

NDI partnered with the TFD for the GCTF event and identified an opportunity to combine the conference with a more hands-on training event for civic groups from across the region. Following the GCTF, NDI organized and led a one-day workshop "Defending Electoral Integrity Against Disinformation," attended by 13 NDI-funded civil society participants, mostly representing citizen election observer groups and fact checking organizations in Asian countries with upcoming elections, as well as guests from the Taiwanese civic tech movement. Building on the information presented during the GCTF, the workshop explored social media monitoring in greater depth. The workshop shared strategies and tools for assessing information environments, navigating the social-media platforms, collecting and analyzing social-media data, developing approaches for countering anti-democratic speech, and holding various stakeholders accountable. This workshop provided citizen observer groups and fact-checkers from the same country the opportunity to work together on mutual support, advocacy, and coordination approaches leading up to their respective elections.

Related efforts have been designed to better build consensus among a broader universe of actors for international observers. For instance, the Carter Center has developed a partnership with grassroots journalism organization Hacks Hackers to conduct a series of workshops among

international election observer groups, other electoral assistance practitioners, international fact-checking networks, academics, and technologists to strengthen interventions and best practices for verifying elections in the face of misinformation and disinformation on social media.

EXPOSING DISINFORMATION THROUGH ELECTION MONITORING

6. KNOWLEDGE-SHARING AND DEVELOPING BEST PRACTICES AROUND COMBATING DISINFORMATION IN ELECTIONS (/TOPICS/MONITORING/6-KNOWLEDGE-SHARING-AND-DEVELOPING-BEST-PRACTICES-AROUND-COMBATING-DISINFORMATION)

In addition to building new partnerships to confront the challenge of disinformation in elections, pre-existing election networks, such as the [Global Network for Domestic Election Monitors](https://gndem.org/) (<https://gndem.org/>) (GNDEM) or the Declaration of Principles for International Election Observation community, can elevate the issue of disinformation, build consensus around defining the challenges that it poses to electoral integrity, and develop best practices to counter it.

As more election monitoring organizations begin to incorporate disinformation monitoring into their broader observation efforts, there are abundant opportunities for peer-to-peer learning and improvement of monitoring methodologies. In September 2019 in Belgrade, Serbia, [NDI conducted an intensive academy for citizen observers from 20 different organizations from around the world on detecting, exposing, and countering malign disinformation](https://gndem.org/stories/gndem-members-convene-in-belgrade-to-discuss-disinformation-in-elections/) (<https://gndem.org/stories/gndem-members-convene-in-belgrade-to-discuss-disinformation-in-elections/>). Participants in the academy learned how disinformation affects electoral integrity, undermines democratic principles, and weakens citizens' trust in elections. The participants shared strategies and methods to monitor disinformation in their own contexts. They walked through exercises on assessing information environments in their countries and practiced using various tools for tracking and analyzing disinformation online. The academy structure encouraged participants to share their organizations' experiences and highlighted lessons learned from working with various social media monitoring tools. For example, ISFED and the CDD-West Africa facilitated discussions and presented on the methods and tools their organizations used to monitor disinformation in their respective contexts.

Participants also explored methods for advocating for greater transparency in online platforms and elevating fact-based political discourse. This included working together to identify ways to hold institutions accountable, build advocacy networks, and create effective messaging to thwart toxic narratives, rooted in each group's local experience.

Knowledge-sharing initiatives have resulted in concrete guidance documents and resources. Over a series of meetings and drafting consultations in the spring of 2019, a small working group representing a mix of international election observers, including NDI, citizen election monitors, academics, fact-checking groups, and civic technologists developed a guide for social media monitoring by civil society, spearheaded by Democracy Reporting International (DRI). [This guide \(https://democracy-reporting.org/dri_publications/guide-for-civil-society-on-monitoring-social-media-during-elections/\)](https://democracy-reporting.org/dri_publications/guide-for-civil-society-on-monitoring-social-media-during-elections/) includes sections on methodology, legal considerations, and tools for social media monitoring in elections by civic groups, working towards creating collective standards and best practices for groups working in the space.

Similar efforts are underway in the international election observation community as part of the continued implementation of the Declaration of Principles. A working group under the DoP is currently building consensus around a framework to observe and assess online campaigns and recommendations grounded in international standards and best practices. As mentioned in the previous section on international election observation, many participating organizations have already begun incorporating this work into their observation missions. The working group presents a chance to identify a set of approaches, rooted in international standards (freedom of expression, transparency, right to political participation, right to privacy, equality and freedom from discrimination, effective remedy) and respective mandates of endorsers of the DoP to assess online campaigns and to seek agreement on a common set of guidelines for the observation of online campaigns by international election observation missions. These guiding principles will be reviewed and endorsed at the DoP annual implementation meeting in Brussels in Spring 2021.

EXPOSING DISINFORMATION THROUGH ELECTION MONITORING

7. CHALLENGES AND ONGOING CONSIDERATIONS FOR MONITORING DIGITAL THREATS IN ELECTIONS (/TOPICS/MONITORING/7-CHALLENGES-AND-ONGOING-CONSIDERATIONS-MONITORING-DIGITAL-THREATS-ELECTIONS)

Unfortunately, with technological advances, digital disinformation efforts and computational propaganda present new and unique challenges to election observation. Identifying networks and connections around the creation, spread, and amplification of disinformation and hate speech in elections is particularly challenging. Online sources lack transparency, with content often spread via fake media houses, phony websites, or social media accounts animated by “farms” of hired users and boosted by automated “bot” accounts.

This is compounded by the fact that the popularity of certain social media platforms and messaging applications varies dramatically by country and platform, as does access to underlying data, while disinformation techniques and content are constantly evolving. The growing popularity of closed messaging services present serious ethical considerations for election observers monitoring their influence. Moreover, the attention, on-the-ground engagement, and implementation of new transparency and content moderation measures provided by online platforms remain inconsistent across national lines. Therefore, monitoring tools and methodologies that may be effective in one context may be irrelevant in another.

The incorporation of social media and other forms of online observation into electoral assessments is in an experimental phase, and monitors are still confronting nascent challenges and identifying lessons learned. These include new technical and political factors that can complicate observations, which may require flexible methodologies to build a more inclusive and comprehensive election assessment.

OBSERVING CLOSED MESSAGING SERVICES

In many countries, campaigning, voter education, and general political discourse around elections is moving to closed messaging services like WhatsApp or Telegram. These networks create serious challenges in terms of what is acceptable to monitor and how to monitor them. Even private channels on public networks (such as closed Facebook Groups) create serious ethical

considerations for any potential study of disinformation. Researchers can consider declaring that they are joining closed groups, as the research group at CDD-West Africa followed in their study. This has the potential, however, to change the nature of conversation within those groups. Another solution is to invite users already in closed groups to submit examples of problematic content, though this approach introduces selection bias and provides an extremely limited view of the closed portion of the online environment. Some civic activists (using tactics that CEPPS does not endorse) have exposed insidious closed syndicates such as hate groups through impersonation or fabricated accounts. This approach violates the terms of services of the platforms and presents serious ethical questions for researchers. Observers must wrestle with these issues to identify an appropriate way to monitor closed platforms, in addition to other methodological challenges, particularly as observers play a different role than traditional academic researchers. Michael Baldassaro, the Digital Threats Lead for the Carter Center, notes: "We do need some consideration that takes into account the law, and ethical considerations that are different from what academic standards might be. I'm not comfortable with going into a WhatsApp group and saying I'm here as a researcher. So, we need to develop modalities for what is appropriate to monitor...and how do we do that?"¹³

EXPOSING ONLINE BARRIERS TO WOMEN AND MARGINALIZED GROUPS TO THE ELECTORAL PROCESS

Information disorder often disproportionately impacts women and marginalized populations as both contestants and voters, often further disadvantaging female candidates and fomenting unsafe online spaces where women and marginalized groups are dissuaded from participating in – or are altogether forced out of - the political discourse. Additionally, many content moderation systems, whether driven by machine learning and artificial intelligence or by direct oversight from human actors, are gender-blind and poorly versed in the local context, including the patterns and dimensions of socio-cultural norms and vulnerabilities of marginalized populations.

However domestic and international organizations surveyed in this research noted that this was an area of concern but not one that they generally addressed specific resources to evaluate. In some cases, the methods, units of analysis, and tools for monitoring hate speech or violence against women online may differ from the broader social media monitoring methodology. For instance, hate speech monitoring may be driven by lexicons of dangerous language, as explored in the methodology developed by NDI and its partners and set forth in [Tweets that Chill: Examining Online Violence Against Women in Politics](https://www.ndi.org/tweets-that-chill) (<https://www.ndi.org/tweets-that-chill>), which rely on examining key words and content. Election monitors may need to balance multiple approaches to derive a real picture of the electoral information landscape and how it affects particular groups. Observer groups should hire gender experts to examine these issues to better understand how existing gender norms function in the local disinformation context, as well as coordinate with groups focused on the impact of disinformation on women and marginalized groups in elections and other critical political contexts. International and domestic observer groups should review

their own implicit bias and cultures of masculinity that can hinder inclusive election observation (/node/13/), particularly as the online space presents new threats to women and marginalized individuals and can reinforce regressive norms.

Hernandez, the Communications Director of MOE, noted that in past missions they had not focused on this in any systematic way, but were interested in developing this capacity in future, and noted groups such as *Chicas Poderosas* that had successfully integrated monitoring for hate speech in recent elections in Brazil, Colombia and Mexico. In Colombia, *Chicas Poderosas* (/interventions/el-poder-de-elegir-de-chicas-poderosas) developed workshops to train local researchers and activists to track hateful political speech on closed messaging groups ahead of the 2018 Presidential Elections.¹⁴ Methodologies such as these to study the content, networks and impact of disinformation and hate speech targeting women and marginalized groups should be more broadly and systematically integrated into election monitoring projects going forward.

NAVIGATING INTERVENTIONS BY SOCIAL MEDIA PLATFORMS

Social media and other technology companies are increasingly responding to the threats that occur on their platforms. In some cases, this has meant providing more transparency about political advertisements on their platforms, more information about group moderators or pages, enhanced responsiveness to flagged content, and specific policies related to managing content that can undermine electoral integrity. However, how and where these initiatives are applied varies drastically from country to country and lacks the level of granularity necessary for robust analysis. In addition, many platforms lack representatives and content moderation in smaller contexts and in countries outside of their major markets. It can be a challenge for observers to gain information about whether, when, and how platforms will respond to any single election. This hampers the ability of observers to develop cogent observation strategies that involve those platforms. Monitoring groups should advocate for enhanced transparency from platforms and work to maintain open lines of communication with these companies, particularly around elections, to enhance corporate accountability and responsibility for safeguarding the online election environment.

DEVELOPING APPROPRIATE AND CONTEXT-SPECIFIC METHODOLOGIES

Variations in how and where citizens consume election information and the dynamic nature of digital threats around elections means there is no “one size fits all” monitoring methodology. Domestic and international groups should consider innovative ways of partnering with each other as well as with fact checkers and advocates for political inclusion of marginalized populations, in order to gain greater insight into the contexts. Social media monitoring may feel overwhelming in scale and scope for election observer groups, with almost limitless numbers of pages, profiles, channels, and volumes of data to potentially collect and analyze. To manage, observers should

develop objectives that are clear, realistic, narrow in scope, and which are derived from a preliminary assessment of the information environment. Subsequent methodologies should seek to achieve these objectives. Only after discrete areas for observation are clarified should groups begin to identify relevant tools that fit the needs of the project and the organization's technical and human resources. In addition, groups should be transparent about the limits of their data and be thoughtful when drawing conclusions.

Observers must consider a range of potential approaches to understand the online election environment. The information age presents new opportunities for developing research to understand how conversations flow online, as well as new challenges to electoral integrity, as trends in discourse are hidden from view in ways that were not possible when the majority of conversations were carried out in traditional media. This is a dynamic and important time for the field to consider the implications of its work in the online space, including the examples and practices analyzed and presented here. Continuous discussion and knowledge exchanges, online and off, will form a key element to countering disinformation through election monitoring. The ability to engage with non-traditional partners such as tech platforms, fact checkers, and others in elections is also crucial. With these considerations reviewed here, observers will be more prepared to address the online environment and integrate it into their planning and recommendations for elections going forward.

DEVELOPING NORMS AND STANDARDS ON DISINFORMATION

0. OVERVIEW - NORMS (/TOPICS/NORMS/0-OVERVIEW-NORMS)

Written by Daniel Arnaudo, Advisor for Information Strategies at the National Democratic Institute

Normative frameworks for the information space have developed over the course of many years, through collaborations between civil society groups, private sector companies, government, and other stakeholders. However, norms and standards specific to working on disinformation or social media issues are in embryonic stages: either existing initiatives are being revised to address the new online threats, for instance through content moderation, corporate governance, the digital agenda, and the cybersecurity space, or new ones dedicated specifically to disinformation and related social media issues are just forming.

This section will examine how the different codes and principles in this space are evolving and how they can potentially link with existing best practices internationally, as well as ways that programs can be designed to link with these nascent frameworks. Some codes work organizationally, for instance how parties, private or public sector entities should behave to

discourage the use and promotion of disinformation, computational propaganda, and other harmful forms of content while encouraging openness, freedom of expression, transparency, and other positive principles related to the integrity of the information space. Others work in terms of individual codes of practice such as for media monitors, fact-checkers, and researchers in the space. Both organizational and individual efforts will be considered in this section.

One way of understanding these normative frameworks for the information space is as a form of negotiation. For example, negotiation between technology companies and other groups (such as governments, advertisers, media, and communications professionals) in agreement on shared norms and standards across non-governmental organizations, media, and civil society that provide oversight and to a certain extent have powers of enforcement of these rules. Different stakeholders enter into different forms of agreement with the information technology and communications sectors depending on the issue agreed on, the principles involved, the means of oversight and safeguards, and ultimately the consequences of any abrogation or divergence from the terms. These standards also focus on the different vectors of information disorder, content, sources, and users. For example, content moderation normative standards such as the [Santa Clara Principles \(https://santaclaraprinciples.org/\)](https://santaclaraprinciples.org/), fact-checking principles focusing on both sources and content by the Poynter Institute's [International Fact-Checking Network \(https://www.poynter.org/ifcn/\)](https://www.poynter.org/ifcn/), or standards such as the [EU Code on Disinformation \(https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation\)](https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation) that attempt to address all three: content through encouraging better moderation, sources by encouraging efforts to identify them, and users through media information literacy standards.

Other actors, such as parties, policymakers, and the public sector, can work to ensure that norms related to online operations are enforced, with varying degrees of success. Ultimately, these normative frameworks are dependent on agreements between parties to abide by them, but other forms of oversight and enforcement are available to society. Also, the integration of inclusive gender-sensitive approaches to the development of norms and standards and reflecting how work to advance gender equality and social inclusion broadly and work to counter disinformation can and should be mutually reinforcing. Many of the frameworks address corporate stakeholders and the technology sector in particular, such as the Santa Clara Principles on Content Moderation, Ranking Digital Rights, and the Global Network Initiative, and the European Union's Codes of Practice on Disinformation and Hate Speech, while others engage with a broader range of groups, including civil society actors, government, media, and communications sectors. Other frameworks attempt to engage with parties themselves, to create codes of online conduct for candidates and campaigns, either through informal agreements or more explicit codes of conduct. Finally, normative frameworks can be used to ensure that actors working in fields related to disinformation issues promote information integrity, such as journalists and fact-checkers.

This section will cover these categories of normative interventions that address content, actors such as platforms, and the targets of disinformation, hate speech, computational propaganda, and other harmful forms of content, including:

- Related multistakeholder norms for cybersecurity, internet freedom, and governance issues (/topics/norms/1-related-multistakeholder-norms-cybersecurity-internet-freedom-and-governance-issues)
 - These standards represent norms for the online space that have impacts on disinformation and related content issues but were not specifically or solely designed to address them. These include the Global Network Initiative, The Manilla Principles on Intermediary Liability, and the Santa Clara Principles.
- Developing codes on disinformation, hate speech, and computational propaganda issues for the private sector (/topics/norms/2-developing-codes-disinformation-hate-speech-and-computational-propaganda-issues)
 - Codes and other normative standards designed specifically to address disinformation and related information integrity issues. These include the EU Codes of Practice on Disinformation and Hate Speech, Ranking Digital Rights, the Global Internet Forum to Counter Terrorism (GIFCT), and the Paris Call for Trust and Security in Cyberspace.
- Party commitments to nonuse of disinformation and computational propaganda and promotion of information integrity principles (/topics/norms/3-party-commitments-nonuse-disinformation-and-computational-propaganda-and-promotion)
 - Standards for parties and individual candidates committing them to information integrity principles, including German Party Commitments, Argentina's Ethical Digital Commitment, Brazil's #NãoValeTudo campaign, Nigeria's Abuja Accord, and the Transatlantic Commission on Election Integrity's Pledge for Election Integrity.
- Codes of conduct for researchers, fact-checkers, journalists, media monitors, and others (/topics/norms/4-codes-conduct-researchers-fact-checkers-journalists-media-monitors-and-others)
 - Broader codes of conduct for those working in the information space such as the Poynter Institute's International Fact-Checking Network's Code of Principles, the Pro-Truth Pledge, Trust Project, Journalism Trust Initiative, and the Certified Content Coalition.

These frameworks all have elements that impact the information space, particularly around freedom of expression, privacy, and the inherent conflicts in creating open spaces for online conversation while also ensuring inclusion and penalties for hateful or other problematic content. They are also evolving and being adapted to the new challenges of an increasingly online, networked society that is confronted by disinformation, hate speech, and other harmful content. This guide will now review more detailed information and analysis of these approaches and potential models, as well as partner organizations, funders, and organizational mechanisms.

DEVELOPING NORMS AND STANDARDS ON DISINFORMATION

1. RELATED MULTISTAKEHOLDER NORMS FOR CYBERSECURITY, INTERNET FREEDOM, AND GOVERNANCE ISSUES (/TOPICS/NORMS/1-RELATED-MULTISTAKEHOLDER-NORMS-CYBERSECURITY-INTERNET-FREEDOM-AND-GOVERNANCE-ISSUES)

Many normative frameworks have developed to govern the online space, addressing issues related to traditional human rights concepts such as freedom of expression, privacy, and good governance. Some of these connect with building normative standards for the online space around disinformation to help promote information integrity but address different aspects of the Internet, technology, and network governance. The Global Network Initiative (GNI) is an older example, which formed in 2008 after two years of development, in an effort to encourage technology companies to respect the freedom of expression and privacy rights of users. The components link with information integrity principles, first by ensuring that the public sphere is open for freedom of expression, secondly by ensuring that user data is protected and not misused by malicious actors potentially to target them with disinformation, computational propaganda, or other forms of harmful content.

The GNI also serves as a mechanism for collective action among civil society organizations and other stakeholders in advocating for better-informed regulation around Information Communication Technologies (ICTs), including social media, to promote principles of freedom of expression and privacy. This includes advisory networks such as the Christchurch Call Network and Freedom Online Coalition, as well as participation in multi-sectoral, international bodies, focused on the issues related to online extremism and digital rights, such as those sponsored by the [United Nations](https://www.un.org/sc/ctc/) (<https://www.un.org/sc/ctc/>) and [Council of](#)



HIGHLIGHT

The [GNI Principles](https://globalnetworkinitiative.org/gni-principles/) (<https://globalnetworkinitiative.org/gni-principles/>), centered around concepts including freedom of expression, privacy, governance, accountability, and transparency, provide a framework for companies to apply human rights principles to their practices, while the [Implementation Guidelines](#)

Europe

(<https://globalnetworkinitiative.org/implementation-guidelines/>) serve as a mechanism for them to be applied in responding to government censorship and surveillance demands.

(<https://www.coe.int/en/web/cybercrime/-/council-of-europe-cooperation-with-internet-sector-two-new-partners>).

REGION	BACKGROUND
GLOBAL	<p><u>The Global Network Initiative</u> (https://counteringdisinformation.org/interventions/global-network-initiative) is an international coalition that seeks to harness collaboration with the technology companies to support The GNI Principles (“the Principles”) and Implementation Guidelines that provide an evolving framework for responsible company decision-making in support of freedom of expression and privacy rights. As our company participation expands, the Principles are taking root as the global standard for human rights in the ICT sector. The GNI also collectively advocates governments and international institutions for laws and policies that promote and protect freedom of expression and privacy for instance through instruments such as the International Covenant on Civil and Political Rights, and subsequently, the United Nations Guiding Principles on Business and Human Rights. It has assessed companies including Facebook, Google, LinkedIn, and Microsoft.</p> <p>GNI Principles:</p> <ul style="list-style-type: none">• Freedom of Expression• Privacy• Responsible Company Decision Making• Multi-Stakeholder Collaboration• Governance, Accountability, and Transparency

In October 2008, representatives of technology companies, civil society, socially responsible investors, and academia released the Global Network Initiative. After two years of discussions, they released a set of principles (<https://globalnetworkinitiative.org/gni-principles/>) focused primarily on how companies that manage Internet technologies could ensure freedom of expression and privacy on their networks. They also established guidelines (<https://globalnetworkinitiative.org/implementation-guidelines/>) for the implementation of these principles. Tech companies with assets related to disinformation, social media, and the overall

information space include Facebook, Google, and Microsoft. Representatives from civil society (<https://globalnetworkinitiative.org/#home-menu>) include the Center for Democracy and Technology, Internews, and Human Rights Watch, as well as representatives from the Global South such as the Colombian Karisma Foundation, and the Center for Internet and Society in India.

Every two years, the GNI publishes an assessment of the companies engaged in the initiative, gauging their adherence to the principles and their success in implementing aspects of them. The latest version was published in April 2020 (<https://globalnetworkinitiative.org/wp-content/uploads/2020/04/2018-2019-PAR.pdf>), covering 2018 and 2019. The principles related to freedom of expression are related to disinformation issues but focus more on companies allowing for freedom of expression rather than preventing the potential harms that come from malicious forms of content such as disinformation and hate speech.

These standards and the GNI have encouraged greater interaction between tech companies and representatives from academia, media, and civil society, and greater consultation on issues related to information integrity, particularly censorship and content moderation. For instance, a Fake News law in Brazil would require "traceability" of users, or registration with government documents within Facebook and other social networks wishing to operate in the country, so that they can be identified for sanction in the case that they are spreading disinformation. This would conflict with the GNI's privacy provisions that ensure users are allowed anonymous access to networks. The GNI released a statement calling out these issues (<https://globalnetworkinitiative.org/gni-concerns-brazil-fake-news-law/>) and has advocated against the proposed law. This shows how this framework can be used for joint advocacy through a multi-stakeholder effort, although its efficacy is less clear. Nonetheless, the GNI has helped form a foundation for other efforts that have since developed, including the Santa Clara Principles on Content Moderation and the EU Codes on Disinformation and Hate Speech that have focused more specifically on social media issues.

Other groups have focused on developing standards linking human rights and other online norms with democratic principles. The Luminate Group's Digital Democracy Charter (<https://luminategroup.com/storage/275/Digital-Democracy-Charter.pdf>), for example, created a list of rights and responsibilities for the digital media environment and politics. The DDC "seeks to build stronger societies through a reform agenda -- remove, reduce, signal, audit, privacy, compete, secure, educate, and inform." In a similar vein, the National Democratic Institute, supported in part by the CEPPS partners, has developed the Democratic Principles for the Information Space, which aim partly to address digital rights issues and counter harmful speech online through democratic standards for platform policies, content moderation, and products.

REGION**BACKGROUND****GLOBAL**[The Manilla Principles on Intermediary Liability](https://counteringdisinformation.org/interventions/manila-principles-intermediary-liability)

(<https://counteringdisinformation.org/interventions/manila-principles-intermediary-liability>)

- Define various principles for intermediary companies to follow when operating in democratic and authoritarian environments, including that: Intermediaries should be shielded from liability for third-party content; Content must not be required to be restricted without an order by a judicial authority; Requests for restrictions of content must be clear, be unambiguous, and follow due process; Laws and content restriction orders and practices must comply with the tests of necessity and proportionality
- Laws and content restriction policies and practices must respect due process; Transparency and accountability must be built into laws and content restriction policies and practices.

[The Manilla Principles on Intermediary Liability](https://www.manilaprinciples.org/principles) (<https://www.manilaprinciples.org/principles>) were developed in 2014 by a group of organizations and experts focused on technology policy and law from around the world. Principle drafters include the Electronic Frontier Foundation, the Center for Internet and Society from India, KICTANET (Kenya), Derechos Digitales (Chile), and Open Net (South Korea) representing a wide range of technology perspectives and regions. They relate to questions of liability for content on networks that have arisen in the US and Europe around Section 230 of the Communications Decency Act of 1996 or Germany's Network Enforcement Act (NetzDG) of 2017.

Manilla Principles on Intermediary Liability

1 Intermediaries should be shielded from liability for third-party content

2 Content must not be required to be restricted without an order by a judicial authority

3 Requests for restrictions of content must be clear, be unambiguous, and follow due process

4 Laws and content restriction orders and practices must comply with the tests of necessity and proportionality

5 Laws and content restriction policies and practices must respect due process

6 Transparency and accountability must be built into laws and content restriction policies and practices

They agreed upon basic standards holding that intermediaries like Facebook, Google, and Twitter, that host content or manage it in some way, should abide by basic democratic standards, while governments should also respect certain norms regarding regulations and other forms of control of content and networks. Their manifesto (<https://www.manilaprinciples.org/principles>) stated:

"All communication over the Internet is facilitated by intermediaries such as Internet access providers, social networks, and search engines. The policies governing the legal liability of intermediaries for the content of these communications have an impact on users' rights, including freedom of expression, freedom of association, and the right to privacy. With the aim of protecting freedom of expression and creating an enabling environment for innovation, which balances the needs of governments and other stakeholders, civil society groups from around the world have come together to propose this framework of baseline safeguards and best practices. These are based on international human rights instruments and other international legal frameworks.

Their principles follow, holding that intermediaries should have legal mechanisms that shield them from liability for the content that they host on their servers. This principle serves to provide for an open conversation and manageable systems of moderation. Secondly, in this vein, the principles assert that content should not be easily restricted without judicial orders, and these must be clear and follow due process. Thirdly, these orders and related practices should comply with tests for necessity and proportionality, or they should be reasonably necessary and

proportional to the gravity of the crime or mistake. Finally, transparency and accountability for these laws should be built into any of these legal systems, so that all can see how they operate and are being applied.

These systems and principles have provided a way for the signatories and other civil society organizations to evaluate how countries are managing online systems, and how platforms can manage their content and apply democratic norms to their own practices. Various organizations have signed on, ranging from media NGOs and organizations, human rights and policy groups, as well as civic technologists. This technical and geographic diversity gives these principles the backing and links to content creators, policymakers, providers, and infrastructure managers, from all over the world (<https://www.manilaprinciples.org/organization-signatories>). They provide one practical means for organizations to work together to monitor and manage these policies and systems related to the information space and in certain cases lobby for changes in them.

"These principles were developed in the wake of a conference at Santa Clara University in 2018. At Santa Clara in 2018, we held the first-of-its-kind conference on content moderation at scale. Most [companies] had not disclosed at all what they were doing. Their policies were about content moderation and how they were applying them. So we co-organized the day-long conference and ahead of this conference a small subgroup of academics and activists organized by the Electronic Frontier Foundation met separately and had a whole conversation and it was out of that sort of side meeting that the Santa Clara principles arose." - Irina Racu, Director of the Internet Ethics Program at Santa Clara's Center for Applied Ethics¹

REGION**BACKGROUND**

[The Santa Clara Principles On Transparency and Accountability in Content Moderation](https://counteringdisinformation.org/interventions/santa-clara-principles-transparency-and-accountability-content-moderation) (<https://counteringdisinformation.org/interventions/santa-clara-principles-transparency-and-accountability-content-moderation>) cover various aspects of content moderation, developed by legal scholars and technologists based mostly in the United States, targeting social media companies with large user bases. The principles include that:

GLOBAL

- Companies should publish the numbers of posts removed and accounts permanently or temporarily suspended due to violations of their content guidelines
- Companies should provide notice to each user whose content is taken down or account is suspended about the reason for the removal or suspension.
- Companies should provide a meaningful opportunity for timely appeal of any content removal or account suspension.

The [Santa Clara Principles On Transparency and Accountability in Content Moderation](https://santaclaraprinciples.org/) (<https://santaclaraprinciples.org/>) developed as a means of assessing how companies are working to develop policies and systems governing the systems that keep track and organize the content that flows on them. Generally, they focus on ensuring that companies have policies that publicize the number of posts removed and accounts banned, provide notice to users when that is done, and provide systems for appeal. Irina Racu, the Director of the Internet Ethics Program at Santa Clara's Center for Applied Ethics, was one of the founders of the project and is a continuing member. She describes how it began:

"Once drafted, various companies signed on in support of them, including social media giants such as Facebook, Instagram, Reddit, and Twitter."

The principles are organized around three overarching themes: **Numbers, Notice and Appeal**. Under **numbers**, platforms agree that companies should keep track and inform the public on the numbers of posts that are reported and accounts that are suspended, blocked, or flagged in a regular report that is machine-readable. Secondly, in terms of the **notice**, users and others who are impacted by these policies should be notified of these takedowns or other forms of content moderation in open and transparent ways. These rules should be published and understood publicly by all users, regardless of background. If governments are involved, say to request a takedown, users should be apprised as well, but generally, those who report and manage these systems should have their anonymity maintained. Thirdly, there should be clearly defined processes of **appeal** for these decisions in place. Appeals should be reviewed and managed by humans, not machines, suggesting mechanisms that groups like the [Facebook oversight board](#)

(<https://staging.counteringdisinformation.org/topics/platforms/0-introduction-platforms>) will attempt to build. However, the principles hold that these practices should be built into all content moderation, not only high-level systems.

These principles have been applied in various ways to draw attention to how companies have developed content moderation systems. One notable application has been the Electronic Frontier Foundation's "Who Has Your Back (<https://www.eff.org/wp/who-has-your-back-2019#santa-clara-principles>)" reports. These reports, released annually, rate companies on the basis of their adherence to the Santa Clara Principles while rating them directly on other metrics as well, such as transparency and notice to users. In their report, EFF notes that 12 of the 16 companies rated in 2019 endorsed the principles (<https://www.eff.org/wp/who-has-your-back-2019#santa-clara-principles>), suggesting that there is some buy-in for the concept. Companies like Reddit adhere to all of the principles, while others like Facebook or Twitter achieve only two or three. With many social media companies still falling short, and international or other new players entering the market, it will remain a challenging effort to apply globally.

DEVELOPING NORMS AND STANDARDS ON DISINFORMATION

2. DEVELOPING CODES ON DISINFORMATION, HATE SPEECH, AND COMPUTATIONAL PROPAGANDA ISSUES FOR THE PRIVATE SECTOR (/TOPICS/NORMS/2-DEVELOPING-CODES-DISINFORMATION-HATE-SPEECH-AND-COMPUTATIONAL-PROPAGANDA-ISSUES)

As demonstrated by these preexisting examples, the private sector is one of the central components of the information ecosystem and has some internal guidelines and norms regulating how it is run. However, there are important normative frameworks that have both induced and encouraged compliance with global human rights and democratic frameworks, and specifically code focused on disinformation, hate speech, and related issues.

The companies that run large platforms in the information ecosystem, such as Facebook, Google, and Twitter, have a special responsibility for the internet's management and curation. There are certain normative frameworks, particularly within the European Union, that governments and civil society have developed to monitor, engage with, and potentially sanction tech companies. Their

efficacy is based on a number of factors, including enforcement and oversight mechanisms in addition to more general threats from harmful media or general adherence to global human rights standards.

The European Union is an important factor as it is a transnational body that has the power to define the conditions to operate in its market. This creates a greater incentive for companies to engage in cooperative frameworks with other private and public sectors as well as civil society actors in negotiation over their rights to operate on the continent. There is the implicit threat of regulation, for instance, the General Data Protection Regulation provides strong data protection that includes not only European citizens but also foreigners who are operating in the country or engaging in systems that are based within it. This implicit power to regulate ultimately provides a significant amount of normative and regulatory pressure on companies to comply if they want to engage in the European common market.

This system creates powerful incentives and mechanisms for alignment with national law and transnational norms. These codes create some of the most powerful normative systems for enforcement around disinformation content, actors, and subjects anywhere in the world but have been challenged by difficulties in oversight and enforcement, while many of the principles would not be permissible in the U.S., particularly concerning potential first amendment infringements. The harmonization of these approaches internationally represents a key challenge in coming years, as various countries impose their own rules on the networks, platforms, and systems, influencing and contradicting each other.

REGION

BACKGROUND

The European Union

(<https://counteringdisinformation.org/interventions/eu-code-practice-disinformation>) developed a Code of Practice on Disinformation (<https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation>) based on the findings of its High-Level Working Group on the issue. This included recommendations for companies operating in the EU, suggestions for developing media literacy programs for members responding to the issues, and developing technology supporting the code.

The five central pillars of the code are:

EUROPEAN
UNION

- enhance the transparency of online news, involving an adequate and privacy-compliant sharing of data about the systems that enable their circulation online;
- promote media and information literacy to counter disinformation and help users navigate the digital media environment;
- develop tools for empowering users and journalists to tackle disinformation and foster a positive engagement with fast-evolving information technologies;
- safeguard the diversity and sustainability of the European news media ecosystem, and
- promote continued research on the impact of disinformation in Europe to evaluate the measures taken by different actors and constantly adjust the necessary responses.

The European Union's Code of Practice on Disinformation is one of the more multinational and well-resourced initiatives in practice currently, as it has the support of the entire bloc and of its member governments behind its framework. The Code was developed by a European Commission-mandated working group on disinformation (<https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation>) and contains recommendations for companies and other organizations that want to operate in the European Union. In addition to the Code, the EU provides member governments and countries that want to trade and work with the bloc with guidelines on how to organize their companies online, as well as plan for responses to disinformation through digital literacy, fact-checking, media, and support for civil society, among other interventions.

The Code was formulated and informed chiefly by the European High-Level Expert Group on Fake News and Online Disinformation in March 2018. The group, composed of representatives from academia, civil society, media, and technology sectors, composed a report that included five

central recommendations that later became the five pillars (<https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation>) under which Code is organized. They are:

1. enhance the transparency of online news, involving an adequate and privacy-compliant sharing of data about the systems that enable their circulation online;
2. promote media and information literacy to counter disinformation and help users navigate the digital media environment;
3. develop tools for empowering users and journalists to tackle disinformation and foster a positive engagement with fast-evolving information technologies;
4. safeguard the diversity and sustainability of the European news media ecosystem, and
5. promote continued research on the impact of disinformation in Europe to evaluate the measures taken by different actors and constantly adjust the necessary responses.

These principles were integrated into the Code, published in October 2018, roughly six months after the publication of the expert group's report. The European Union invited technology companies to sign on to the Code and many engaged, alongside other civil society stakeholders and EU institutions that worked to implement elements of these principles. Signatories included Facebook, Google, Microsoft, Mozilla, Twitter, as well as the European Association of Communication Agencies, and diverse communications and ad agencies. These groups committed not only to the principles, but to a series of annual reports on their progress in applying them, whether as communications professionals, advertising companies, or technology companies.

As participants in the initiative, the companies agree to a set of voluntary standards aimed at combating the spread of damaging fakes and falsehoods online and submit annual reports on their policies, products, and other initiatives to conform with its guidelines.

(<https://techcrunch.com/2020/06/22/tiktok-joins-the-eus-code-of-practice-on-disinformation/>). The initiative has been a modest success in engaging platforms in dialogue with the EU around these issues and addressing them with members governments, other private sector actors, and citizens.

The annual reports of these companies and the overall assessment of the implementation of the Code of Practice on Disinformation review the progress that the code has made in its first year of existence, from October 2018-2019. The reports find that while the Code has generally made progress in imbuing certain aspects of its five central principles in the private sector signatories, it has been limited by its "self-regulatory nature, the lack of uniformity of implementation and the lack of clarity around its scope and some of the key concepts."

An assessment from September 2020 found that the code had made modest progress but had fallen short in several ways, and provided recommendations for improvement. It notes that "[t]he information and findings set out in this assessment will support the Commission's reflections on pertinent policy initiatives, including the European Democracy Action, as well as the Digital Services Act, which will aim to fix overarching rules applicable to all information society services." This helps describe how the Code on Disinformation fits within a larger program of European

initiatives, linking with similar codes on hate speech moderation, related efforts to ensure user privacy, copyright protection, and cybersecurity, and broader efforts to promote democratic principles in the online space.

Other organizations have made independent assessments that offer their own perspective on the European Commission's project. The project commissioned a consulting firm, Valdani, Vicari, and Associates (VVA), to review the project as well, and it found that:

- "The Code of Practice should not be abandoned. It has established a common framework to tackle disinformation, its aims and activities are highly relevant and it has produced positive results. It constitutes a first and crucial step in the fight against disinformation and shows European leadership on an issue that is international in nature.
- Some drawbacks related to its self-regulatory nature, the lack of uniformity of implementation and the lack of clarity around its scope and some of the key concepts.
- The implementation of the Code should continue and its effectiveness could be strengthened by agreeing on terminology and definitions."

The Carnegie Endowment for International Peace completed an assessment in a similar period after the completion of its first year of implementation, published in March 2020 (<https://carnegieendowment.org/2020/03/03/eu-code-of-practice-on-disinformation-briefing-note-for-new-european-commission-pub-81187>). The author found that the EU had indeed made progress in areas such as media and information literacy, where several technology signatories have created programs for users on these concepts, such as Facebook, Google, and Twitter.

The EU Code of Practice on Disinformation's normative framework follows similar, related examples that describe and develop a component of the European Union's position, namely the 2016 EU Code of Conduct on Countering Illegal Hate Speech. This 2016 EU Code of Conduct links with the earlier "Framework Decision 2008/913/JHA (<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM%3A133178>) of 28 November 2008 combating certain forms and expressions of racism and xenophobia by means of criminal law" and national laws transposing it, means all conduct publicly inciting to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, color, religion, descent or national or ethnic origin." Alternatively, organizations such as the Center for Democracy and Technology have criticized the EU's approach and potential for misuse and abuse, particularly in regards to the code on hate speech. (<https://cdt.org/insights/letter-to-european-commissioner-on-code-of-conduct-for-illegal-hate-speech-online/>)

Overall, both the European Commission and Carnegie reports found that there is much still to be done and that the Code on Disinformation would benefit from better-shared terminology and structure. To that end, the EU recently adopted its Democracy Action Plan (https://ec.europa.eu/info/files/communication-european-democracy-action-plan_en). Countering disinformation is one of its core pillars, with the effort to improve the EU's existing tools and impose costs on perpetrators, especially on election interference; to move from Code of Practice

to a co-regulatory framework of obligations and accountability of online platforms consistent with the Digital Services Act; and to set up a framework for monitoring the implementation of the code of practice.

As can be seen, while companies have signed onto the EU Codes on Disinformation and Hate Speech, and member governments have pledged to follow their principles, oversight, and enforcement are separate, more difficult mechanisms to apply. Nonetheless, with the force of other countries, in other regions, these codes or similar kinds of agreements could provide a framework for collaboration around various issues related to disinformation, hate speech, online violent extremism, and a host of other harmful forms of content.

REGION	BACKGROUND
GLOBAL	<p><u>Ranking Digital Rights Normative Frameworks</u> (https://counteringdisinformation.org/interventions/ranking-digital-rights)</p> <p>Ranking Digital Rights (RDR) ranks the world's most powerful digital platforms and telecommunications companies on relevant commitments and policies, based on international human rights standards.</p> <p>The RDR principles focus on three central pillars: Governance, Freedom of Expression, and Privacy.</p>

For many years, technologists, academics, and other civil society representatives have worked together to push the private sector to address digital rights issues. An example is the [Ranking Digital Rights](https://rankingdigitalrights.org/) (<https://rankingdigitalrights.org/>), an initiative sponsored by the New America Foundation that focuses on creating a concrete framework to engage companies around normative issues related to the information space. Starting in 2015, Ranking Digital Rights has published a "[Corporate Accountability Index](https://rankingdigitalrights.org/index2020/) (<https://rankingdigitalrights.org/index2020/>)" that ranks technology, telecom, and Internet companies on their commitments to human rights. This framework is rooted in international human rights principles such as the Universal Declaration of Human Rights (UDHR) and [United Nations Guiding Principles on Business and Human Rights](https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf) (https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf).

The indicators cover principles related to governance, freedom of expression, and privacy and give companies a score based on their compliance with various aspects of the Index. Companies that are ranked by the Index include major players in social media, search, and other issues related to the information space including Facebook, Google, Microsoft, and Twitter. Their responsiveness to these principles provides indications of how initiatives either inspired by or analogous to Ranking

Digital Rights can address social media, hate speech, and disinformation issues, while linking to older initiatives around corporate accountability that preceded it, such as the Global Network Initiative.

Rebecca MacKinnon, a former journalist and digital rights scholar, board member of the Committee to Protect Journalists, and a founding member of the Global Network Initiative, created the Ranking Digital Rights project (RDR) in 2013 partly based on her book, *Consent of the Networked*. Nathalie Marechal, a Senior Policy Analyst at the project, details how the book was "one of the first pieces of research that honed in on the role that the private sector plays and tech companies specifically play in human rights violations both when they act as agents of governments as a result of government demands for data or demands for censorship, and as a result of companies pursuing their own business interests. The book ended with a call to action to push companies for transparency and more accountability for their role in enabling or perpetrating human rights violations."

The RDR principles focus on three central pillars: Governance, Freedom of Expression, and Privacy. From these central principles, the project developed indicators that serve to measure and evaluate a company's adherence to these core tenets. These were developed to apply not only to what they call "mobile and internet ecosystems" companies, but also telecommunications companies such as Verizon or T-Mobile. It divides its surveys into these two categories and assigns the companies scores out of 100 based on their compliance and adherence to the indicators under the principles. These scores are tabulated and combined into a final score that is explored in *Indexes* (<https://rankingdigitalrights.org/index2019/>), which are backed by data and were published semi-annually from 2015 up until 2019, with a new edition due in 2021.

The indexes are somewhat dynamic in that they evolve based on new technologies or developments in the field, as well as new scholarship, which has changed the categories that define the methodology, the indicators, and the companies reviewed. For instance, the mobile and internet ecosystem was known simply as the Internet in 2015 and renamed *internet and mobile in 2017* (<https://rankingdigitalrights.org/2017-indicators/>). The RDR project publishes the *methodology* (<https://rankingdigitalrights.org/index2019/report/index-methodology>) openly and allows for others to *adapt it under creative commons license* (<https://rankingdigitalrights.org/adaptations/>) to produce their own ratings, for instance for local or national companies. As a result, the RDR system has been replicated in contexts such as *India* (<https://cis-india.org/internet-governance/files/ranking-digital-rights-in-india.pdf>), *the Middle East* (<https://smex.org/dependent-yet-disenfranchised-the-policy-void-that-threatens-the-rights-of-mobile-users-in-arab-states/>), and *Africa* (https://www.accessnow.org/cms/assets/uploads/2018/02/RDR-Africa_Final-version-5_January-2018.pdf).

This is part of a process the organization has developed to keep the principles relevant while also stable enough to provide data about how companies are improving or declining in terms of the index. This has helped to develop and expand the index to focus on 24 companies including telcos like AT&T and Telefónica as well as social media platforms and tech companies like Facebook, Google, Microsoft, and Twitter. This summary gives a general view of the RDR system and the

areas and indicators it covers. It touches on information space issues in various ways and includes major technology companies with purview over a large scale, global social media networks, such as Facebook, Google, Microsoft, and Twitter. Within this system, they also consider properties these companies control, such as WhatsApp (Facebook) or Skype (Microsoft). These companies generally score similarly on the indicators, earning overall scores of 62 (Microsoft), 61 (Google), 57 (Facebook), and 55 (Twitter). By contrast, Chinese and Russian telecom companies score much lower, such as the Chinese tech giant Tencent (home to WeChat, QQ, and QZone) at 26, the search engine and tech services goliath Baidu at 23, or the Russian Yandex at 32. This certainly serves to contrast the approaches of companies in both authoritarian and democratic spheres of influence, and the contrast on human rights grounds that can be useful to emphasize, especially with regards to increasingly prevalent information integrity and disinformation issues.

RDR Governance Indicators

G1. Policy commitment

G2. Governance and management oversight

G3. Internal implementation

G4. Impact assessment

G5. Stakeholder engagement

G6. Remedy

Under governance, the principles look for ways that a tech corporation governs itself and its products. This connects with the way that they manage their platforms, what kind of oversight they have in place, and particularly how they assess the impact that these platforms are having. As they note in their 2019 Index Report: "Indicator G4 evaluates if companies conduct risk assessments to evaluate and address the potential adverse impact of their business operations on users' human rights. We expect companies to carry out credible and comprehensive due diligence to assess and manage risks related to how their products or services may impact users' freedom of expression and privacy." This is increasingly becoming a key component of companies' policies concerning disinformation issues, and to how they can govern themselves effectively with regards to human rights concerns around freedom of expression and privacy issues in particular.

The Index also notes how no company, including platforms like Facebook, Google, and Twitter, are making assessments about the impact of artificial intelligence or ways to "identify and manage the possible adverse effects of rules enforcement on users' freedom of expression and privacy rights," nor risk assessments of the human rights implications of the design and implementation of their

terms of service or targeted advertising systems. These internal public company policies are having huge impacts on the information environment, and RDR provides one means of evaluating them.

RDR Freedom of Expression Indicators

F1. Access to terms of service

F2. Changes to terms of service

F3. Process for terms of service enforcement

F4. Data about terms of service enforcement

F5. Process for responding to third-party requests for content or account restriction

F6. Data about government requests for content or account restriction

F7. Data about private requests for content or account restriction

F8. User notification about content and account restriction

F9. Network management (telecommunications companies)

F10. Network shutdown (telecommunications companies)

F11. Identity policy

The freedom of expression indicators relates more specifically to the governance of the content in online platforms that are being evaluated. The terms of service help define the way that companies determine users' rights in access, complaints, suspension, and takedown processes.

RDR evaluates how they have made information about these terms and changes to them available to users, and then secondarily provides publicly available information about the process through which takedowns or restrictions on content are made, as well as overall data about the kinds of takedowns there are. This also relates to the ways that governments make take-down requests and notes that Facebook (<https://transparency.facebook.com/>), Google (<https://transparencyreport.google.com/?hl=en>), and Twitter (<https://transparency.twitter.com/>) have all been making more data available about take-downs through transparency reports, except

for government request-related data, which has become more limited. Facebook and Twitter have been releasing less data related to government requests for data, particularly in the case of requests on closed platforms like Facebook Messenger, WhatsApp, and Twitter's Periscope video platform (<https://staging.counterinformation.org/topics/platforms/0-introduction-platforms>).

It also looks at company policies around identity, if companies require users to provide government-issued ID or some other form of identification that could be tied to their real-world identity. This could allow for better identification of sources of disinformation and hate speech, or other nefarious users, but also creates potential avenues for targeting vulnerable users by governments, trolls, and others. They note that Google, Instagram, WhatsApp, and Twitter allow anonymous users across their platforms, but that Facebook requires identification, something that can create conflicting problems, particularly for vulnerable users.

RDR Privacy Indicators

- P1. Access to privacy policies**
- P2. Changes to privacy policies**
- P3. Collection of user information**
- P4. Sharing of user information**
- P5. The purpose for collecting and sharing user information**
- P6. Retention of user information**
- P7. Users' control over their own user information**
- P8. Users' access to their own user information**
- P9. Collection of user information from third parties
(internet companies)**
- P10. Process for responding to third-party requests for user
information**
- P11. Data about third-party requests for user information**
- P12. User notification about third-party requests for user
information**
- P13. Security oversight**
- P14. Addressing security vulnerabilities**
- P15. Data breaches**
- P16. Encryption of user communication and private content
(internet, software, and device companies)**
- P17. Account Security (internet, software, and device
companies)**
- P18. Inform and educate users about potential risks**

Finally, in terms of privacy issues, RDR covers how different policies related to user data and information about how it is handled, how its security is ensured, how vulnerabilities are addressed, and how oversight and notification about breaches are addressed. While these issues may seem tangential to disinformation campaigns, they can actually have major impacts, as data that is taken from these companies can often be used in disinformation campaigns, users that are accessing content through weak security systems can be spied on by governments and other

nefarious actors, and targets of disinformation campaigns or cyber-attacks may be unaware that they are even under attack without the proper systems for monitoring that their access is secure or to be notified in cases of breach. They also examine if companies inform users about potential "cyber risks," which they define (<https://rankingdigitalrights.org/2019-indicators/#cyberrisks>) as "[s]ituations in which a user's security, privacy, or other related rights might be threatened by a malicious actor (including but not limited to criminals, insiders, or nation-states) who may gain unauthorized access to user data using hacking, phishing, or other deceptive techniques." This could include risks from targeted, online disinformation or harassment campaigns, particularly for vulnerable or marginalized users.

As a component of its ongoing review of tech practices and policies, RDR is evolving to examine issues around the ethical use of private data and algorithms to provide content. The 2020 Index, will include considerations of these issues based on its revision. It has already been revised over a period of several years to cover evolving information systems, such as mobile phones, social media, and other technologies.

As Marechal notes: "We kept the methodology steady between 2017-2018 and for 2019 there were a couple of tweaks and we added companies every year, but by-and-large we kept it comparable for those three research cycles and there was measurable progress for most companies across the years in mid-2018. We started a project to revise and expand the RDR methodology and that was a project that I led, to account for human rights harms associated with two interrelated issues, business models based on targeted advertising and the use of algorithms (<https://rankingdigitalrights.org/its-the-business-model/>). The use of what our funder calls it called AI and that we called algorithmic systems in consumer-facing products focusing specifically on their use for Content moderation and content governance." They have also translated the methodology into other languages, including Arabic, French, and Spanish. (<https://rankingdigitalrights.org/translations/>) This provides a further basis to internationalize and localize the framework for various contexts globally.

REGION	BACKGROUND
GLOBAL	Global Internet Forum to Counter Terrorism (GIFCT) (https://counteringdisinformation.org/interventions/global-internet-forum-counter-terrorism) fosters collaboration and information-sharing between the technology industry, government, civil society, and academia to counter terrorist and violent extremist activity online.

Terrorist organizations and individual actors have carried out attacks against civilians and critical infrastructure to instill fear, chaos, and reduce both geopolitical and internal cohesion of societies for a long time. Since the introduction of the internet and, most especially, social media, terrorist organizations have used the web to radicalize individuals, gain supporters, the technical "know-

how” about building bombs and improvised explosive devices, and spread disinformation and propaganda to populations. What’s particularly noteworthy in recent years is the power of and the use of social media platforms by terrorist organizations. The 2019 Christchurch New Zealand Shooting, where the video of the shooter was initially posted on Twitch but reshared on YouTube, Facebook, and Twitter, provides a prime example of terrorists’ use of technology and the internet to spread their narratives and disinformation.

In response to increased terrorist activity in the information environment, the Global Internet Forum for Counter-Terrorism (<https://gifct.org/>)(GIFCT) was formally established in 2017 by 4 core companies: Twitter, Microsoft, Facebook, and YouTube, as well as several smaller signatories that increased its reach across platforms. GIFCT has been designed to foster collaboration and information-sharing between industry partners to thwart terrorist actors’ ability to use the information environment to manipulate, radicalize, and exploit targeted populations. The four companies that made up the forum took turns in chairing the work of GIFCT. Following the Christchurch call to strengthen the coordinated response to terrorism in cyberspace through a multistakeholder process, GIFCT has become its own non-profit organization and is currently managed by its first inaugural Executive Director, Nicholas Rasmussen, former Director of the National Counterterrorism Center. The goals of GIFCT are:

- Improve the capacity of a broad range of technology companies, independently and collectively, to prevent and respond to abuse of their digital platforms by terrorists and violent extremists.
- Enable multi-stakeholder engagement around terrorist and violent extremist misuse of the internet and encourage stakeholders to meet key commitments consistent with the GIFCT mission.
- Encourage those dedicated to online civil dialogue and empower efforts to direct positive alternatives to the messages of terrorists and violent extremists.
- Advance broad understanding of terrorist and violent extremist operations and their evolution, including the intersection of online and offline activities.

A core aspect of GIFCT is knowledge sharing and cooperation, not only with the main tech platforms but with smaller ones as well. As such, GIFCT is working with Tech Against Terrorism (<https://www.techagainstterrorism.org/>), a private-public partnership launched by the UN Counter-Terrorism Executive Directorate (UN CTED). The goals of this effort are to provide resources and guidance to increase knowledge sharing within the tech industry; encourage peer learning and support amongst members; foster collaboration and information sharing between the tech sector, government, civil society, and academia; and promote greater understanding about ways that terrorists exploit the internet to achieve their objectives.

PARIS CALL FOR TRUST AND SECURITY IN CYBERSPACE

With the rise of both disinformation campaigns and cyberattacks in cyberspace, and the shared understanding of the need for increased collaboration and cooperation to foster technological innovation yet prevent attacks in cyberspace, a group of 78 countries, 29 public authorities, 349 organizations, and 648 companies have come together to align around a set of nine principles to create an open, secure, safe, and peaceful cyberspace. The Paris Call reaffirms these countries with the commitment to international humanitarian and customary international law that provides the same protections for citizens online the way these laws apply offline. In creating this call, governments, civil society, and industry, including social media companies, adhere to providing safety, stability, and security in cyberspace, as well as increased trust and transparency to citizens. The call has created a multi-stakeholder forum process for organizations and countries to come together to increase information sharing and collaboration. Participants to the Paris Call have signed onto the following nine principles:

1. Prevent and recover from malicious cyber activities that threaten or cause significant, indiscriminate, or systemic harm to individuals and critical infrastructure.
2. Prevent activity that intentionally and substantially damages the general availability or integrity of the public core of the Internet.
3. Strengthen our capacity to prevent malign interference by foreign actors aimed at undermining electoral processes through malicious cyber activities.
4. Prevent ICT-enabled theft of intellectual property, including trade secrets or other confidential business information, with the intent of providing competitive advantages to companies or to the commercial sector.
5. Develop ways to prevent the proliferation of malicious software and practices intended to cause harm.
6. Strengthen the security of digital processes, products, and services, throughout their lifecycle and supply chain.
7. Support efforts to strengthen advanced cyber hygiene for all actors.
8. Take steps to prevent non-State actors, including the private sector, from hacking-back, for their own purposes or those of other non-State actors.
9. Promote the widespread acceptance and implementation of international norms of responsible behavior as well as confidence-building measures in cyberspace.

These principles have been signed onto by states such as Colombia, South Korea, and the UK, although not the United States initially, CSOs including IRI, IFES, and NDI; private sectors such as telecom (BT), social media (Facebook), and information technologies (Cisco, Microsoft); as well as a host of other companies. The Call provides a framework for normative standards related to cybersecurity and disinformation across sectors, particularly under the third principle focused on building capacity to resist malign influence in elections.

DEVELOPING NORMS AND STANDARDS ON DISINFORMATION

3. PARTY COMMITMENTS TO NONUSE OF DISINFORMATION AND COMPUTATIONAL PROPAGANDA AND PROMOTION OF INFORMATION INTEGRITY PRINCIPLES (/TOPICS/NORMS/3-PARTY-COMMITMENTS-NONUSE-DISINFORMATION-AND-COMPUTATIONAL-PROPAGANDA-AND-PROMOTION)

Parties are a critical component of political systems, and their adherence to normative frameworks is a challenging but central part of any political system's susceptibility to disinformation and other negative forms of content. When candidates and parties adhere to normative standards, for instance, to refrain from the use of computational propaganda methods and the promotion of false narratives, it can have a positive effect on the integrity of information in political systems. When parties, particularly major players in political systems, refrain from endorsing these standards or actively work to adopt and adapt such kinds of misleading methods, such as disinformation campaigns and computational propaganda, this can have an incredibly harmful effect on the kind of content being promoted and the potential for false narratives, conspiracies, and hateful and violence-inducing speech to permeate and dominate campaigns. It is worth examining examples of parties working together to create positive standards for the information environment, as well as interventions for encouraging this kind of environment.

In the first system, parties can develop their own codes, either individually or collectively. One of the better examples of this is the German political parties during the 2017 parliamentary campaign season. Other than the right-wing Alliance for Germany (Afd) party, all of the parties agreed to the non-use of computational propaganda, the spread and endorsement of false narratives, and other tactics. Germany has a regulatory framework in the social media space, linked with EU regulations such as the Global Data Protection Regulation, which provides useful data privacy for European citizens as well as those who simply access European networks.

In other cases, civil society can work together to induce parties to develop and adhere to codes of practice on disinformation, hate speech, and other information integrity issues. In Brazil, various civil society groups came together in the 2018 election to develop a public code of norms for parties and candidates to follow. The NãoValeTudo campaign tried to encourage politicians to

adopt the motto that "Not everything is acceptable" (*Não Vale Tudo*), which included not promoting false content, not engaging in false networks or the automating of accounts for false purposes, and other norms to ensure that the campaigns were acting fairly and in line with principles that would encourage an open and fair conversation about policy and society. This was formed by a consortium of groups including fact-checking groups like Aos Fatos, digital rights organizations such as Internet Lab and the Institute of Technology and Equity, and the national association of communications professionals (*Associação Brasileira das Agências de Comunicação – ABRACOM*).

COUNTRY	BACKGROUND
BRAZIL	<p>#NãoValeTudo (/interventions/nao-vale-tudo) (https://staging.counteringdisinformation.org/interventions/nao-vale-tudo)(Not everything is acceptable) is a code of ethics for politicians, civic groups, and parties to follow that was developed during the 2018 Brazilian election cycle. The code focuses on principles around the non-use of computational propaganda techniques such as bot or troll networks, the non-promotion of false claims, transparency around campaign use and non-abuse of private user data, and the promotion of a free and open information space. Politicians and parties could signal their support through social media posts tagging the phrase, which was supported by a wide coalition of CSOs.</p>

The group declared that:

"recent examples concern us, as they indicate that activities such as the collection and misuse of personal data to target advertising, the use of robots and fake profiles to simulate political movements, and positions and methods of disseminating false information can have significant effects on rights of access to information, freedom of expression and association, and privacy of all and all of us. The protection of such rights seems to us to be a premise for technology to be a lever for political discussion and not a threat to the autonomy of citizens to debate about their future."

The group received some endorsements, most notably from presidential candidate Marina De Silva the former Minister of Environment for former President Lula De Silva's past government, and a relatively high-level



HIGHLIGHT

candidate, who put out social media on her adherence, encouraging others to join

NãoValeTudo outlined principles that campaigns, politicians, and other organizations could adhere to, including:

We need to know how we are using technology in politics and to take collective responsibility for the consequences of these uses.

We do not tolerate the production and dissemination of false news. Whoever creates them, promotes lies and manipulates citizens around private and dishonest interests.

We believe that detailed information on the use of technologies for electoral purposes should be public knowledge, such as software, applications, technological infrastructure, data analysis services, professionals, and companies involved in the construction and consultancy of our campaign.

We reject the manipulation of the public's perception of the political discussion carried out from the creation and use of false profiles.

The use of bots, however, can be beneficial for the construction of political debates, but the use of these tools must always be ostensibly informed because robots that impersonate humans can be a great obstacle to a transparent, open, collective debate. plural and constructive.

We defend freedom of expression and criticism of citizens in the electoral period.

We believe that data is valuable and important in campaigns to enhance the dialogue between candidates and citizens, but that its use must be carried out

responsibly.

(https://www.facebook.com/marinasilva.official/posts/1900415863303368?comment_id=2089072314661294&comment_tracking=%7B%22tn%22%3A%22R%22%7D). While other local candidates also endorsed them, they did not receive buy-in from others in the presidential race, including the eventual winner, Jair Bolsonaro. Nonetheless, they created a platform for discussion of disinformation issues and the acceptability of certain online tactics in the online sphere through the #NãoValeTudo hashtag and other methods, while also raising general awareness of these threats and highlighting how reluctant many campaigns and politicians were to embrace them. This methodology could be replicated by other civil society groups to develop standards for parties, call out those who break the rules, and raise awareness among the general public.

In a third form, international coalitions have worked together to form normative frameworks. Ahead of the 2019 Argentine Elections, in cooperation with Argentina's Council on Foreign Relations (CARI: Consejo Argentino para las Relaciones Internacionales) and organized by the National Electoral Council (CNE: Cámara Nacional Electoral), the Woodrow Wilson International Center for Scholars, the Annenberg Foundation, and International IDEA developed an Ethical Digital Commitment "with the aim of avoiding the dissemination of fake news and other mechanisms of disinformation that may negatively affect the elections.." Hosted by the CNE, parties; representatives of Google, Facebook, Twitter, and WhatsApp; organizations of media, and internet and technology professionals signed this Commitment. Parties and other organizations would help to both implement and provide oversight for it. These approaches show practical, often multisectoral, approaches and collaboration between public, private, and political sectors, in addition to civil society, on these issues, following similar efforts by election management bodies in Indonesia and South Africa, as explained in the [EMB section \(/topics/embs/0-overview-emb-approaches\)](/topics/embs/0-overview-emb-approaches).

Similar, earlier codes have focused on hateful or dangerous speech in addition to other elections-related commitments, such as agreeing to accept a result. One such example developed in Nigeria ahead of its 2015 elections is how the presidential candidates pledged to avoid violent or inciting speech in the so-called "[Abuja](#)

(<https://www.idea.int/sites/default/files/codesofconduct/Abuja%20Accord%20January%202015.pdf>).

(<https://www.idea.int/sites/default/files/codesofconduct/Abuja%20Accord%20January%202015.pdf>).

developed with support from the international community and former UN Secretary-General Kofi Annan. This represented a particular effort to protect the rights of marginalized groups to participate in the electoral process and "to refrain from campaigns that will involve religious incitement, ethnic or tribal profiling both by ourselves and by all agents acting in our names." In an effort more focused on information integrity itself, the Transatlantic Commission on Election Integrity, a group made up of a "bi-partisan group of political, tech, business and media leaders", developed The Pledge for Election Integrity for [candidates of any country](#).

(<https://www.electionpledge.org/>) to sign. Its principles (<https://www.electoral.gob.ar/nuevo/paginas/pdf/CompromisoEticoDigital.pdf>) are outlined in the highlight box to the right.

The pledge has gained over 170 signatories in Europe, Canada, and the United States, and also has the potential to expand to other contexts. A commission named for the late Kofi Annan, former head of the UN, also endorsed the pledge, suggesting that it could be translated for other contexts: "We endorse the call by the Transnational Commission on Election Integrity for political candidates, parties, and groups to sign pledges to reject deceptive digital campaign practices. Such practices include the use of stolen data or materials, the use of manipulated imagery such as shallow fakes, deep fakes, and deep nudes, the production, use, or spread of falsified or fabricated materials, and collusion with foreign governments and their agents who seek to manipulate the election." Nonetheless, with any of these pledges there remain challenges of enforcement and wide-ranging acceptance among political candidates, especially in polarized or deeply contested environments. Standards development in this area remains a challenge, but a potentially critical mechanism for building trust in candidates, parties, and overall democratic political systems.



HIGHLIGHT

The Pledge for Election Integrity

Committing not to fabricate, use or spread data or materials that were falsified, fabricated, doxed, or stolen for disinformation or propaganda purposes;

Avoiding dissemination, doctored audios/videos or images that impersonate other candidates, including deep fake videos;

Making transparent the use of bot networks to disseminate messages; avoid using these networks to attack opponents or using third-parties or proxies to undertake such actions;

Taking active steps to maintain cybersecurity and to train campaign staff in media literacy and risk awareness to recognize and prevent attacks;

Committing to transparency about the sources of campaign finances.

DEVELOPING NORMS AND

STANDARDS ON DISINFORMATION

4. CODES OF CONDUCT FOR RESEARCHERS, FACT CHECKERS, JOURNALISTS, MEDIA MONITORS, AND OTHERS (/TOPICS/NORMS/4-CODES-CONDUCT-RESEARCHERS-FACT-CHECKERS-JOURNALISTS-MEDIA-MONITORS-AND-OTHERS)

COUNTRY	BACKGROUND
GLOBAL	The Poynter Institute's International Fact-Checking Network (/interventions/international-fact-checking-network) has developed a Code of Principles for fact-checkers to follow globally that includes standards around the methodology of the practice. Groups are vetted to ensure that they follow the standards and those that are found to be in compliance are admitted to the network. The network has become the basis for Facebook's fact-checking initiative, among others that have proliferated globally in contexts ranging from the EU and US to countries across the global south.

Fact-checking and other forms of research are generally described in our section on [civil society](https://staging.counterindisinformation.org/topics/csos/0-introduction-building-civil-society-capacity) (<https://staging.counterindisinformation.org/topics/csos/0-introduction-building-civil-society-capacity>) but the concept is derived from key normative frameworks in research and ethical mechanisms for building trust in industries, communities, and society as a whole. The Poynter Center's International Fact Checking Network (IFCN) is a network of newspapers, television, media groups, and civil society organizations that are certified by the IFCN to review content in ways that conform with international best practices. This is basically ensuring that the process and standards for fact-checking follow honest, unbiased guidelines and certify that the organizations and their staff understand and comply with these rules. IFCN standards link with earlier journalistic standards to source, develop, and publish stories, such as the [Journalist's Creed](https://en.wikipedia.org/wiki/Journalist%27s_Creed) (https://en.wikipedia.org/wiki/Journalist%27s_Creed), or national standards of [journalism associations](https://www.spj.org/ethicscode.asp) (<https://www.spj.org/ethicscode.asp>).

Other standards have been developed specifically for journalists, such as The Trust Project (<https://thetrustproject.org/>), funded by Craig Newmark Philanthropies. The Trust Project designed a system of indicators about news organizations and journalists in order to ensure reliable information for the public and encourage trust in journalism. These have been created to create norms that media organizations and social media can follow in order to maintain a standard of information released. This group has partnered with Google, Facebook, and Ring to "use the Trust Indicators in display and behind the scenes," according to their website, and has been endorsed by over 200 news organizations such as the BBC, El Pais, and the South China Morning Post. This project has also been translated and replicated in contexts such as Brazil (<https://staging.counterindisinformation.org/interventions/projeto-credibilidade>), and invites journalists and organizations from around the world to join.

In a similar project, Reporters Without Borders, the Global Editors Network, the European Broadcasting Union, and Agence France Presse have formed a similar Journalism Trust Initiative (<https://www.journalismtrustinitiative.org/>) (JTI) to create similar standards for journalism ethics and trustworthiness. The initiative "is a collaborative standard setting process according to the guidelines of CEN, the European Committee for Standardization" according to its explanation of its history and process on the JTI's website (<https://www.journalismtrustinitiative.org/>). Also funded by Newmark, through a multiyear, multistakeholder process to develop and validate standards starting in 2018, the JTI seeks to build norms among journalists, promoting compliance within the community of news-writing, particularly to combat mistrust in journalism and disinformation.

Poynter International Fact-Checking Network Standards

- *A Commitment to Nonpartisanship and Fairness*
- *A Commitment to Transparency of Sources*
- *A Commitment to Transparency of Funding and Organization*
- *A Commitment to Transparency of Methodology*
- *A Commitment to Open and Honest Corrections*

The IFCN standards begin with nonpartisanship and fairness, something that is often difficult to guarantee in ethnically diverse, polarized, or politicized situations. Fact-checking groups must commit to following the same process for any fact check they do, and without bias towards content in terms of source, subject, or author. This ensures the fact-checkers are fair and neutral. They must also be transparent and show their sources and how they arrived at their answer, and this should be replicable and documented, with as much detail as possible. These groups must also be transparent about their funding sources and how they are organized and implement their work. Staff must understand this transparency and work to engage in their business in this way. The methodology that they use must also be presented and practiced in an open way so that anyone can understand or even replicate what the group is publishing. This creates an

understanding of a fair and level system for reviewing and printing judgments about content. Finally, when the group gets something obviously wrong, they must agree to issue rapid and understandable corrections.

Groups take courses and pass tests showing that their systems and staff are cognizant of the standards and implement them in their practice. Groups also publish their standards, methodologies, and organizational and funding information publicly. The head of the IFCN Baybars Orsek described the process:

" Those organizations go through a thorough and rigorous application process involving external assessors and our Advisory Board and in positive cases end up being verified, and platforms...particularly social media companies like Facebook and others often use our certification as a necessary, but not a sufficient criteria to work with fact-checkers right now."

Verified organizations that pass these tests join the network, link with partner organizations, participate in training, collaborate on projects, and work with other clients as trusted fact-checkers, particularly social media companies such as Facebook that have engaged fact-checkers from the network in contexts all over the world. This concept is covered further in the [Platform-Specific Engagement for Information Integrity topical section \(/topics/platforms/0-overview-platforms\)](#), but generally, groups that work with Facebook have their fact checks integrated directly into the Facebook application, allowing for the app itself to show the fact checks next to the content. While this methodology is still being developed, and the efficacy of fact checks continues to be difficult to confirm, it provides a much more visible, dynamic, and powerful system to apply them. Groups that want to join the network can apply and this can help ensure that when a project begins, they have a proper and complete understanding of the state of the field and best practices in terms of fact-checking work.

Certain organizations have tried to expand normative frameworks beyond journalists and fact checkers to broader civil society, with varying degrees of success. The [Certified Content Coalition \(https://credibilitycoalition.org/credcatalog/project/certified-content-coalition/\)](#) had a goal of standardizing requirements for accurate content by bringing together various organizations in support of initiatives for new norms and standards. These groups consist of a research cohort of journalists, students, academics, policy-makers, technologists, and non-specialists interested in the mission of the program. The Certified Content Coalition's goal is to create a widespread understanding of information being disseminated to the public in a way that is collaboratively agreed upon by groups, allowing for a greater sense of credibility. It ultimately stalled, with its founder [Scott Yates noting, \(https://journalist.net/certified-content-coalition-next-chapter\)](#) "[a]dvertisers said they wanted to support it, but in the end it seems that the advertising people were more interested in the perception of doing something than in actually doing something. (In hindsight, not shocking.)" This result potentially highlights the limits of these kinds of initiatives.

The broader **Pro-Truth Pledge** (<https://www.protruthpledge.org/>) is an educational nonpartisan nonprofit organization focused on science-based factual decision making. The pledge is for politicians and citizens to sign to commit to truthful political systems to promote facts and civic engagement. While it has a much wider potential reach, its application and the measurement of its effect is much more challenging. However, as with other norms, it has the potential to raise public awareness around information integrity issues, foster conversation, and potentially grow trust in good information and critical thinking around the bad.

HELPING PARTIES PROTECT THE INTEGRITY OF POLITICAL INFORMATION

0. OVERVIEW - POLITICAL PARTY APPROACHES (/TOPICS/PARTIES/0-INTRODUCTION-INTEGRITY)

Written by Bret Barrowman, Senior Specialist for Research and Evaluation, Evidence and Learning Practice at the International Republican Institute, and Amy Studdart, Senior Advisor for Digital Democracy at the International Republican Institute

CONCEPTUAL FRAMEWORK (/TOPICS/PARTIES/1-POLITICAL-PARTIES-AND-TRAGEDY-INFORMATION-COMMONS)

Even in relatively democratic, competitive political party environments, two related dilemmas make countering disinformation difficult. First, competitive parties face a "tragedy of the commons" (</topics/parties/1-political-parties-and-tragedy-information-commons#tragedy>) with respect to disinformation, in which a healthy information environment leads to the best social outcomes, but also incentivizes individual actors to gain a marginal electoral advantage by muddying the waters. Second, parties are not unitary, but are collections of distinct candidates, members, supporters, or associated interest groups, each with its own interests or incentives. In this case, even when party organizations are committed to information integrity, they face a "principal-agent" (</topics/parties/1-political-parties-and-tragedy-information-commons#principal-agentproblem>) dilemma in monitoring and sanctioning co-partisans. These related dilemmas create an incentive for political parties and candidates to avoid engaging in or implementing

programmatic responses. Democracy, human rights, and governance (DRG) funders and implementing partners can mitigate these dilemmas by using networking and convening power to help parties maintain commitments to information integrity, within and between parties.

PROGRAMMATIC RESPONSES (/TOPICS/PARTIES/2-PROMOTING- INFORMATION-INTEGRITY-MULTI-PARTY- POLITICAL-SYSTEMS)

DRG practitioners have implemented a wide range of programmatic approaches to reduce both the impact and use of disinformation and related tactics by political parties during elections. These approaches are summarized in the table below, according to the “core party function(s)” – the functions that parties perform in an ideal-type democratic party system – upon which the program approach might be expected to operate. This typology is intended to provide DRG practitioners with a tool through which to analyze party systems and programmatic approaches, with the goal of designing programs that are tailored to the challenges of political party partners.

PROGRAM APPROACHES	CORE PARTY FUNCTIONS			
	Interest Articulation (expressing citizen interests through electoral campaigns or implementation of policy)	Interest Aggregation (bundling many disparate, and occasionally conflicting, citizen interests into a single branded policy package or platform)	Mobilization (activating citizens, usually party supporters, for political engagement, including attending rallies or events, taking discrete actions like signing petitions or contacting representatives, and especially voting.	Persuasion (parties' or candidate attempt to change voters', undecided voters or opposition supporter opinions on candidate or policy issues.

PROGRAM APPROACHES	CORE PARTY FUNCTIONS			
Programs on Digital Media Literacy	*	*	*	
Programs on AI and Disinformation			*	*
Programs for Closed Online Spaces and Messaging Apps		*	*	
Programs on Data Harvesting, Ad Tech & Microtargeting		*	*	*
Programs on Disinformation Content and Tactics	*	*		
Research Programs on Disinformation Vulnerability and Resilience	*	*		
Programs for Understanding the Spread of Disinformation Online			*	*
Programs Combating Hate Speech, Incitement, and Polarization		*	*	

PROGRAM APPROACHES	CORE PARTY FUNCTIONS			
Policy Recommendations and Reform/ Sharing and Scaling Good Practice in Programmatic Responses	*	*	*	*

RECOMMENDATIONS (/TOPICS/PARTIES/3-POLICY-RECOMMENDATIONS)

- When implementing these programmatic approaches, consider political incentives in addition to technical solutions.
- Programmatic interventions should account for diverging interests within parties – parties are composed of functionaries, elected officials, interest groups, formal members, supporters, and voters – each of which may have unique incentives to propagate or take advantage of disinformation.
- The collective action problem of disinformation makes one-off interactions with single partners difficult – consider implementing technical programs with regular, ongoing interaction between all relevant parties to increase confidence that competitors are not “cheating.”
- Relatedly, use the convening power of donors or implementing organizations to bring relevant actors to the table.
- Consider pacts or pledges, especially in pre-election periods, in which all major parties commit to mitigating disinformation. Importantly, the agreement itself is cheap talk, but pay careful attention to design of institutions, both within the pact and externally, to monitoring compliance.
- There is limited evidence for effectiveness of common counter-disinformation program approaches with a focus on political parties and political competition, including media literacy, fact-checking, and content labeling. That there is limited evidence does not necessarily imply these programs do not work, only that DRG funders and implementing partners should invest in the rigorous evaluation of these programs to determine their impact on key outcomes like political knowledge, attitudes and beliefs, polarization,

propensity to engage in hate speech or harassment, and political behavior like voting, and to identify what design elements distinguish effective programs from ineffective ones.

- DRG program responses have tended to lag political parties' use of sophisticated technologies like data harvesting, microtargeting, deep fakes and AI generated content. Funders and implementing partners should consider the use of innovation funds to generate concepts for responses to mitigate the potentially harmful effects of these tools, and to rigorously evaluate impact.

HELPING PARTIES PROTECT THE INTEGRITY OF POLITICAL INFORMATION

1. POLITICAL PARTIES AND THE TRAGEDY OF THE INFORMATION COMMONS (/TOPICS/PARTIES/1-POLITICAL-PARTIES-AND-TRAGEDY-INFORMATION-COMMONS)

DEFINITION OF POLITICAL PARTIES

Political parties are organized groups of individuals with similar political ideas or interests who try to make policy by getting candidates elected to office.¹ This electoral function – advancing candidates for office and securing votes for those candidates – distinguishes political parties from other organizations, including civil society organizations (CSOs) or interest groups. This electoral role creates unique incentives for political party actors with respect to disinformation and programmatic responses.

POLITICAL PARTIES, INFORMATION, AND DEMOCRACY: AN OVERVIEW FOR DEVELOPING

CONTEXT ANALYSIS, PROBLEM STATEMENTS, AND THEORIES OF CHANGE

HOW PARTIES CONNECT CITIZENS WITH THEIR REPRESENTATIVES

The ability of party systems to constructively shape electoral competition depends on the exchange of high-quality information. Conceptually, parties connect citizens to elected officials through a market mechanism. In democratic multiparty systems, political parties bundle many disparate, and occasionally conflicting, interests into a single branded package (*interest aggregation*) which they in turn “sell” to voters during elections (*interest articulation*).² Importantly, however, this process represents an ideal model of democratic competition between programmatic political parties that political scientists expect to produce the best democratic outcomes for citizens, including high quality public goods and services, and high levels of accountability. However, no single party or party system approximates this model in practice, and many fall short of it.

Indeed, in many cases, parties fail to effectively aggregate or articulate citizen preferences. Disinformation, creating fractured, isolated epistemic communities, clearly makes the processes of interest aggregation and articulation more difficult, although it is ultimately unclear whether disinformation is a cause or consequence. For these processes to operate effectively, political parties and elected officials must have good information about the preferences of their constituents, and voters must have good information about the performance of their representatives. Party brands facilitate this accountability by providing a yardstick for voters; citizens can judge their representatives against what their party brand promises. These processes are particularly important for political inclusion. Clear information about constituent preferences and representatives’ performance improves the likelihood that the interests of marginalized groups are heard and perceived as legitimate, and as such, provides an electoral incentive for political leaders to address those interests. This transmission of information between elites and voters is a necessary (but not sufficient condition) for democratic party systems to function. Without good information, parties and elected officials cannot ascertain constituent preferences, and voters cannot associate performance or policy outcomes with a party brand to hold elected officials accountable. Furthermore, disinformation can influence whose voices are heard and what interests are legitimate. As such, political elites may have an incentive to use disinformation to further marginalize under-represented groups.

EXCLUDABILITY AND ATTRIBUTION: WHY IT IS HARD FOR CITIZENS TO HOLD REPRESENTATIVES ACCOUNTABLE FOR PUBLIC POLICIES WITHOUT FUNCTIONING PARTIES AND GOOD INFORMATION.

However, this problem of the exchange of good information is compounded by the nature of public policies. In economic terms, public goods and services are *non-excludable* – it is difficult to prevent individual citizens from enjoying them if they are provided. For example, a good national defense establishment protects all citizens, even those who have not paid their taxes; it is not practical or cost effective for a state to withhold national defense from specific citizens. Private goods -- money in exchange for a vote, for example -- can be delivered directly to specific individuals, who know exactly who provided it.

Public policies, on the other hand, suffer from a problem of *attribution*. Since these goods are provided collectively, citizens may be less sure what specific officials or parties are responsible for them (and conversely, who is responsible for unintended consequences or the lack of policy altogether). Also, public policies are complicated. Both these policies, and their observable outcomes for citizens, are the products of complex interactions of interests, context, policymaking processes, and implementation. Furthermore, observable outcomes, like a good economy or a healthy population, may significantly lag the policies that are most directly responsible for them. As such, citizens may find it difficult to attribute policy outcomes to specific representatives.³ Political parties can help simplify complex policy issues for voters, again assuming an exchange of good information between elites and voters.

THE TRAGEDY OF THE INFORMATION COMMONS: ACCOUNTING FOR INCENTIVES IN COUNTERING DISINFORMATION PROGRAMS

These interrelated concepts – the *interest aggregation* and *articulation* functions of parties, role of *information* in democratic political competition, and the *attribution* problems of public policies have important implications for the design and implementation of counter-disinformation programs. Like national defense or a functioning transportation infrastructure, a healthy information environment benefits everyone, and it is impractical to exclude individuals or single groups from that benefit. For parties, this nature of the information environment creates a collective action or “free-rider” problem.⁴ While the best collective outcomes occur when all actors refrain from engaging in disinformation, each individual has an incentive to “free-ride” – to enjoy

the healthy information environment while gaining a marginal competitive advantage by muddying the waters. In this sense, the problem of disinformation for political parties is a tragedy of the commons,⁵ in which small transgressions by multiple actors end up spoiling the information environment.⁶ This can occur even in ideal circumstances – relatively open environments with competitive elections. It is compounded in authoritarian or semi-authoritarian systems in which the incumbent exercises significant control over the information environment through repression or control of media outlets, or where fringe parties or politicians have an incentive to proliferate provocative content with the goal of increased attention or visibility.⁷ This control of the information environment precludes meaningful electoral competition between parties, further reducing any incentive to cooperate on information integrity. While this situation may create incentives for opposition parties to counter disinformation, especially if they see gains from public perceptions of honesty, it may also lead to vicious cycles of degrading the information environment when there are alterations of power.

Like other public goods and services, a good information environment benefits everyone. Citizens get accurate information about how their representatives are doing and can reward or sanction them accordingly. Parties get good information about what their citizens want. A good information environment depends on every actor committing to this outcome. In fact, parties have an electoral incentive to muddy the waters – to let every other competitive party be honest while they misrepresent issues of public policy. Again, this dilemma makes countering disinformation difficult even in the best-case scenario. Where parties and party systems fall short of this ideal type, the dilemma will be more difficult to resolve.

THE PRINCIPAL-AGENT PROBLEM OF POLITICAL PARTIES: MAINTAINING COMMITMENTS TO COUNTERING DISINFORMATION *WITHIN* PARTIES

Furthermore, political parties are not unitary; they are coalitions of varied (and often competing) candidates, constituencies, and interest groups. As such, all political parties face an additional challenge of keeping candidates and members accountable to the



HIGHLIGHT

Principal-Agent Problem: An organizational problem in which one actor (the principal) has authority to set collective goals and must ensure that one or more other actors (the agents) behave in a way that advances those goals, despite the agents controlling information about their own performance. For an illustration of the principal-agent problem in campaign messaging, see Enos, Ryan D., and Eitan D. Hersh. “Party Activists as Campaign Advertisers: The Ground Campaign as a Principal-Agent Problem.”

parties' organizational goals and platform. In the context of disinformation, even democratically inclined or reform parties, or parties that think they can gain votes by taking a stand against disinformation, confront a *principal-agent problem*. On the one hand, party leaders may simply be unaware of affiliates' attempts to generate or take advantage of

disinformation. On the other hand, this problem creates plausible deniability – elites may tacitly encourage supporters to engage in disinformation to help the parties' electoral prospects while the leadership signals a commitment to information integrity. In addition, often, individual party members exploit gender or other identity-based cleavages of “competitors” within their own party to gain a competitive edge, that can include the use of hate speech, disinformation or other harmful forms of content promoted in the public sphere. If this dynamic is unacknowledged, DRG programming can help legitimize campaign tactics that undermine democratic accountability. In short, DRG practitioners should not assume political parties are unitary, and technical solutions should include approaches to helping political party actors ensure that all candidates and supporters maintain commitments to information integrity. While these models help illustrate important incentives that program designers should be sensitive to, it is important to note that they do not preclude technical solutions. Beyond providing encouragement, support, and training for party leadership in setting tone and expectations, establishing infrastructure for communication and coordination within the party will hold members and candidates accountable. The “[DRG Program Responses to Disinformation with Political Party Partners](https://staging.counterdisinformation.org/topics/parties/2-drg-program-responses-disinformation-political-party-partners)” (<https://staging.counterdisinformation.org/topics/parties/2-drg-program-responses-disinformation-political-party-partners>) section below provides concrete ideas for programs to support parties' efforts to protect information integrity.

American Political Science Review 109, no. 2 (May 2015): 252–78.
<https://doi.org/10.1017/S0003055415000064>
(<https://www.google.com/url?q=https://doi.org/10.1017/S00030554150000648>)

PARTY FUNCTIONS, INCENTIVES FOR ABUSE, AND PROGRAM DESIGN

In concrete terms, political parties perform four information-based functions in democratic multi-party systems: *interest aggregation*, *interest articulation*, *citizen mobilization*, and *persuasion*. Democratic collective outcomes are more likely when parties perform these functions based on good information. However, within each function, parties or individual candidates have incentives to manipulate information to gain an electoral advantage.

Interest aggregation refers to parties' capacity to solicit information about citizen interests and preferences. To develop responsive policies and compete in elections, parties must have reliable information about the interests and preferences of the voters. They may also have an incentive to mischaracterize public sentiment both to their opponents and the public. For instance, to

prioritize a policy that is broadly unpopular, but which is important for a key specific constituency, a party or candidate might have an incentive to mischaracterize a public opinion study, or to artificially amplify support for a policy on social media using bot networks.

Interest articulation refers to parties' ability to promote ideas, platforms, and policies, both in the campaign and policymaking process. Interest articulation requires political parties to engage in both mass and targeted communication on issues with voters. This function may also require parties and candidates to *persuade* (see below) citizens of their viewpoints – particularly to convince voters that specific policies will fulfill those voters' interests. Again, there is a social benefit to “true” information about the policies and positions – citizens can cast their votes for the parties that best represent their preferences. However, to gain an electoral advantage, individual parties or candidates may have an interest in misrepresenting their policy positions or the potential consequences of preferred policies, by fabricating research studies or by scapegoating vulnerable groups.

Mobilization refers to parties' capacity to activate citizens for political engagement, including attending rallies or events, taking discrete actions like signing petitions or contacting representatives, and especially voting. To produce the most democratic outcomes, mobilization should be based on good information; parties should provide potential voters with accurate information about policies and the electoral process – particularly where and how to vote. However, individual parties and candidates can gain an electoral advantage by engaging in more nefarious mobilization tactics. Mobilization can involve coercion – the use of disinformation to “scare” voters about the consequences of opponents' policies, to activate voters by inflaming prejudices or political cleavages, or to demobilize opposition candidates or supporters through harassment or by generating apathy.

Persuasion refers to parties' or candidates' attempt to change voters' opinions on candidates or policy issues. In contrast to mobilization, which often focuses on known party supporters or apathetic voters, persuasion is usually targeted to moderates, “undecideds” or weak supporters of opposing parties.

IMPLICATIONS

These interrelated concepts – the *excludability* and *attribution* problems of public policies, the concept of the information space as a *tragedy of the commons*, and the role of information in the *interest aggregation, articulation, and mobilization* functions of political parties have several concrete implications for practitioners designing and implementing counter-disinformation programs:



HIGHLIGHT

Key Agents: Trolls, bots, fake news websites, conspiracy theorists, politicians, partisan media outlets, mainstream media outlets, and foreign governments⁹

1. The best collective democratic outcomes require a healthy information environment.
2. Each individual actor (a party or candidate) may perceive an incentive to let others behave honestly while they try to gain a competitive advantage through disinformation.
3. As such, political parties have an incentive both to perpetuate and to mitigate disinformation. Whether they choose to perpetuate or mitigate depends on the context; in short, parties want voters to have true information about things that help them and false information about things that hurt them. The inverse is true for their political opponents.
4. Information disorders are a product of many actors perceiving this incentive structure and knowing their competition is acting according to these incentives. Parties might be willing to commit to information integrity if they could be confident their competitors would do the same simultaneously. However, if they are not confident their opponents will do the same, **even honest or democratically inclined parties may be unwilling or unable to forgo using disinformation if it means losing elections and their opportunity to implement their agenda.**
5. When political parties ARE committed to information integrity, they are not unitary, and face an additional challenge of keeping candidates, members, and supporters accountable to the parties' commitments. This unwillingness or inability to monitor and sanction co-partisans compounds the dilemma in point four above.
6. Consider these political incentives before designing and implementing technical

Key Messages¹⁰: causing offense, affective polarization, racism/sexism/misogyny, "social proof" (artificial inflation of indicators that a belief is widely held), harassment, deterrence (use of harassment or intimidation to discourage an actor from taking an action, like running for office or advocating for a policy), entertainment, conspiracy theories, fomenting fear or anxiety of a preferred in-group, logical fallacies, misrepresentations of public policy, factually false statements

Key Interpreters: Citizens, party members and supporters, elected officials, members of the media



KEY RESOURCE

NOTES ON SOURCES OF DISINFORMATION:

The information disorder framework (<https://www.coe.int/en/web/freedom-expression/information-disorder>) suggests identifying disinformation tactics and possible responses by thinking systematically about the agent, the message, and the interpreter of the information. Many DRG programs, especially funded by USG donors, focus on building resilience within a target country to foreign disinformation campaigns, especially by the governments (or pro-government supporters) of China, Russia, and Iran. However, it is important to note, especially among political parties, that the

solutions. Frameworks like Thinking and Working Politically and Applied Political Economy Analysis

agents (or perpetrators) of disinformation campaigns may also be domestic actors seeking to affect the behavior of their political opposition or supporters. Even in the case of foreign-directed campaigns, interpreters (or targets) are not selected broadly or arbitrarily. Rather, foreign campaigns seek to exacerbate existing social and political cleavages, with the goal of eroding trust in institutions writ large. Furthermore, foreign campaigns often rely on witting or unwitting supporters in target countries. In both cases of foreign and domestic disinformation campaigns, therefore, historically marginalized groups including women, ethnic, religious, or linguistic minorities, persons with disabilities, and LGBTI individuals are often disproportionately targeted and injured by these efforts.⁸

(<https://www.usaid.gov/documents/1866/thinking-and-working-politically-through-applied-political-economy-analysis>) can help practitioners better understand the unique political incentives facing potential partners and beneficiaries. Key political solutions may include internal and external monitoring and coordination between relevant parties in committing to mitigating disinformation.

A note on technical solutions: In some cases, particularly with democratic or reform parties, parties may express a willingness to take concrete steps to mitigate disinformation. In a smaller set of cases, all major competitive parties might be willing to take these steps. Even these best case scenarios where parties are engaged in building solutions give rise to problems surrounding technical solutions, including resource constraints and technical capacity. However, technical solutions are necessarily secondary to more fundamental political solutions.

HELPING PARTIES PROTECT THE INTEGRITY OF POLITICAL INFORMATION

2. PROMOTING INFORMATION INTEGRITY IN MULTI-PARTY POLITICAL SYSTEMS (/TOPICS/PARTIES/2-PROMOTING-INFORMATION-INTEGRITY-MULTI-PARTY-POLITICAL-SYSTEMS)

Over the last few years, DRG practitioners have implemented a wide range of programmatic approaches to reduce both the impact and use of disinformation and related tactics during elections. Most DRG programmatic approaches look at the overarching information ecosystem, which has incidental impacts for political party behavior and the impact of disinformation as a campaign tactic. However, the last few years have seen an increasing number of interventions targeted specifically at political parties. These programmatic approaches operate on a wide variety of theories of changes, with various implicit or explicit assumptions about incentive structures for political parties, voters, and other electoral information actors. Keeping in mind the relevant functions of political parties, and the potential benefits to disinformation agents, including domestic actors, the following section outlines broad programmatic approaches that have been applied to assist political party partners in building resilience to disinformation. Generally, DRG programs tend to operate from a similar coherent logic – that if partners and/or their voters can identify disinformation and have the technical capacity to deter or respond to it, they will improve the information environment, leading to more responsiveness and accountability.

DIGITAL MEDIA LITERACY PROGRAMS

The European Union's Joint Research Commission definition on digital and information literacy through its [Digital Competence Framework](http://ftp.jrc.es/EURdoc/JRC68116.pdf) (<http://ftp.jrc.es/EURdoc/JRC68116.pdf>) is key to understanding the effects of digital literacy programs. Their definition of digital competency is a Venn diagram of intersecting



HIGHLIGHT

One promising approach to media literacy, for example, was partnership with state educational institutions to implement media literacy at scale. The [IREX Learn2Discern](https://www.irex.org/project/learn-discern-l2d-media-literacy-training) (<https://www.irex.org/project/learn-discern-l2d-media-literacy-training>) campaign

literacies including media, information, internet and ICT, which touches on different aspects of digital competency, from using the internet and understanding information in the abstract, to using ICTs in terms of hardware and software and the media in different forms. All these literacies are important in understanding how programs can address disinformation vulnerability and resilience.

One approach to election related disinformation is to increase public awareness of the what, where, why and how of disinformation. Education campaigns vary in scope - both who they reach and what they address - and can be run by several different actors, including CSOs, schools, faith-based

institutions, technology companies, and governments. The theory of change is that if the electorate is aware of the presence of disinformation and the ways in which it operates, then they will be more critical of the information they encounter and that this will then have less of an impact on their political views. Broadly, this approach is among the likeliest to have positive ramifications outside of election integrity and can - where implemented at scale - reduce the impact of health misinformation and susceptibility to cybercrime.

In this sense, media literacy programs can operate on the "*interest aggregation*" and "*mobilization*" functions of political parties by mitigating the impact of disinformation on polarization, especially among strong partisans.

AI AND DISINFORMATION PROGRAMS

This program approach includes assistance to political party partners in responding to the use of a range of artificial intelligence (AI) applications, including automated artificial amplification, deep fakes, and the manipulation and modification of audio and videos. Increasingly these approaches encompass the use of large networks of automated accounts with more intelligently informed content, shaped by user responses, personal data and other metrics.

Efforts to combat disinformation have, as of writing, largely focused on the human-led creation of misleading content; false amplification using fake accounts, paid followers, and automated bots; and paid promotion of misleading content microtargeted at users based on their probable susceptibility to a given narrative. This focus has mirrored the widespread accessibility and scalability of the technologies underpinning disinformation - bots, content farms, fake followers, and microtargeted ads that have radically changed the way election and political information is

implemented media literacy trainings through community centers, schools, and libraries. Rigorous evaluations of the Learn2Discern program in Ukraine ([/topics/csos/2-fact-checking#Ukrainefactcheck](https://www.rand.org/pubs/research_reports/R3.html)) have found that both youth and adult learners were significantly more likely to be able to identify false news stories from true news, and that short media literacy videos and source labels mitigated the impact of Russian propaganda content (https://www.rand.org/pubs/research_reports/R3.html).

created and distributed. For political parties, these technologies facilitate the social nature of their core functions, particularly by artificially signaling “social proof” – that a proposed policy or candidate is more widely supported than it is. These technologies also game trending and recommendations algorithms, increasing exposure for information (or disinformation) that otherwise may not have been widely available.¹¹ Artificial amplification, therefore, helps parties and candidates manipulate citizen beliefs, rather than responding to constituent interests directly. While we do not expect that there will be any major shifts in the vectors for election-related disinformation, the DRG community is increasingly concerned about the further automation of content creation and distribution and the ease of access to “deep fake” manipulation, in which a video or audio is created of a person saying or doing something they never said or did.

Technological approaches have been developed to identify areas where image and audio have been altered by detecting anomalies in pixelation or audio waves. At present, those are not deployed on a systematic basis. Given that deep fakes and AI-generated content have not yet begun to play a major role in campaigns or election integrity, case studies for DRG programmatic interventions have been limited.

Although DRG programmatic responses have been limited, research organizations have established knowledge repositories on problems of computational propaganda. For example, the Oxford Internet Institute, through the [Program on Computational Propaganda](https://staging.counteractingdisinformation.org/interventions/comprop-navigator) (<https://staging.counteractingdisinformation.org/interventions/comprop-navigator>), developed the ComProp Navigator, a curated collection of resources for civil society organizations to consider when responding to disinformation issues.

PROGRAMS FOR CLOSED ONLINE SPACES AND MESSAGING APPS

Programs countering disinformation on applications that are 'closed' (or private) and encrypted networks must consider the difficulty in accessing user data and the privacy considerations in collecting this data.

Disinformation campaigns are rapidly moving from the relatively public sphere of online social media and content platforms like Facebook, YouTube, and Twitter, to private messaging platforms such as WhatsApp, Line, Telegram, and SMS. Several of those platforms are encrypted, making it a challenge to track and prevent the spread of false content and amplification. In several instances, political parties have exploited private messaging to target supporters who then forward misleading messages to other supporters – giving little opportunity for independent or opposition actors to counteract or correct messaging.

Several programmatic approaches have emerged to combat this challenge. In Taiwan, [a civil society group created an initiative](https://cofacts.g0v.tw/) called “CoFacts (<https://cofacts.g0v.tw/>)”, to address the large scale spread of political disinformation on LINE. Messages can be forwarded to the CoFacts bot for fact checking by a team of volunteers; the CoFacts bot can also be added to private groups and

will automatically share corrections if a fact-checked piece of false content is shared within the group. This preserves the privacy of the group writ large, while allowing for the monitoring and countering of false information.

In several countries with contentious elections or political situations, Facebook (the parent company of WhatsApp) has limited the size of WhatsApp groups and the number of times a message can be forwarded, which reduces the ease and potential for virality on the platform. Another approach is to flood encrypted or private messaging services with accurate information. For example, the Taiwanese government has employed a number of comics and comedians (<https://qz.com/1863931/taiwan-is-using-humor-to-quash-coronavirus-fake-news/>) to create fact-based, easy to forward content designed for virality.

FEATURED INTERVENTION COFACTS (/INTERVENTIONS/COFACTS)

PROGRAMS ON DATA HARVESTING, AD TECH & MICROTARGETING

Programs countering the use of private user data in targeted disinformation campaigns are in their infancy as of this writing. However, these approaches are becoming increasingly important as this user data can be used to inform automated systems and ad buys in political campaigns. Programs include efforts to reverse engineer these systems to illuminate their ubiquity and effect.

Data harvesting, advertising technology, and microtargeting increasingly feature prominently in parties' *mobilization* (</topics/parties/0-introduction-integrity#mobilization>) and *persuasion* (</topics/parties/0-introduction-integrity#persuasion>) functions. Advertising tools allow political parties to tailor messages to small groups based on demographic, attitudinal, behavioral, and geographic characteristics gleaned from a variety of sources, including online behavior. This capacity to tailor political messages to smaller and smaller constituencies has important implications for democratic outcomes. Individual parties and candidates use this technology because it helps optimize their messaging. Socially, however, the adoption of this technology has two important consequences.

First, it undercuts the *interest aggregation* function of political parties. Recall that democratic outcomes are most likely when parties effectively bundle disparate interests and policies under one brand for "sale" to broad swaths of voters. The microtargeted communications facilitated by

advertising technology allow single parties or candidates to tailor messages directly to small groups. This approach may produce short-run gains in mobilization effectiveness at the expense of negotiating common policy priorities and building consensus around issues. Second, the adoption of this technology also facilitates the more precise targeting of disinformation, hate speech, harassment, and other nefarious tactics that parties or candidates might employ to activate their own supporters or suppress the engagement of supporters of their political opponents. Third, microtargeting effectively “hides” content from the media, fact-checkers, or opposing parties who might otherwise be able to respond to or debate the information in question.¹²

DRG programs to encourage best practices in, and discourage abuses of, advertising technology have tended to lag on the adoption of these methods. One example, however, is the Institute of Mass Information (<https://imi.org.ua/en/>) in Ukraine which monitors social media platforms during elections. Their Executive Director noted in 2019 that Facebook was not particularly effective in addressing abusive political advertising, especially, “native” or “sponsored” content – political advertising disguised to look like news. In this case, Facebook’s political advertising database was not useful to third-party monitors because the advertising content was so difficult to detect.¹³ This challenge provides one example of how innovations in advertising technology might undermine democratic outcomes, especially when they provide electoral benefits to individual parties or candidates; if political messages are disguised to look like factual information or news, by precisely targeting to consumers attitudes, tastes, or behaviors, producers are more likely to manipulate citizens’ preferences than respond to them.

PROGRAMS ON DISINFORMATION CONTENT AND TACTICS

Programs examining disinformation content and tactics take on a wide variety of forms, whether simply collecting and analyzing the information or looking to infiltrate disinformation groups to study their methods. These approaches also play an important accountability function with respect to political parties. A focus on the content of disinformation may help citizens and CSOs clarify complex policy issues, reducing the space for parties and candidates to muddy the water. This approach sees either independent journalists, volunteers, or CSOs check the veracity of content, issue corrections, and – in some instances – work with social media companies to flag misleading content, limit its spread, and post the fact-checkers correction alongside a post. Some of these initiatives target political party or candidate content explicitly, while others look at the broader information ecosystem and fact-check stories based on their likely impact, spread, or a specific interest area.

Programs to develop fact-checking and verification outlets are rarely done in direct partnership with political party actors given that the approach requires political neutrality to be effective. However, hypothetically, these programs serve an important accountability function by acting on the incentives of political actors. A theory of change underlying these approaches is that if political

actors, especially elected officials, know that false statements will be identified and corrected in a public forum, they may be less likely to engage in this behavior in the first place. Furthermore, fact-checking and verification outlets can provide accurate information to voters, who may then more effectively punish purveyors of disinformation at the ballot box. In Ukraine, for example, a program funded by the British Embassy and implemented by CASE Ukraine (<https://case-ukraine.com.ua/en/projects/budget-simulator-for-pilot-united-territorial-communities/>) developed a set of information technology (IT) tools to enable citizens to analyze state budgets, in theory to develop critical thinking to counter politicians' populist rhetoric on complex economic issues.¹⁴ Similarly, support for "explainer journalism" modeled on outlets like Vox.com (<https://www.vox.com/>) in the United States has emerged as an approach to counter parties' attempts to confuse citizens on complex policy issues. VoxUkraine (<https://voxukraine.org/en/>), for example, supported by several international donors and implementing partners, provides both fact-checking, explainers, and analytical articles, especially on issues of economic reform in Ukraine.¹⁵

Program approaches have also drawn on pop culture, using satire and humor to encourage critical thinking around disinformation on complex issues. For example, Toronto TV (<https://www.youtube.com/user/uttoronto>), supported by the National Endowment for Democracy, Internews, and Pact, and inspired by American satirical takes on news and current events by Jon Stewart, John Oliver, Hassan Minhaj and others, use social media platforms and short video segments to challenge disinformation narratives propagated by prominent politicians.

A number of the interventions aimed at this issue have focused on countering disinformation ahead of election cycles and understanding the role of social media in spreading information during modern political campaigns, such as International IDEA's roundtable on "Protecting Tunisian Elections (<https://www.idea.int/news-media/news/protecting-tunisian-elections-digital-threats>)," held in 2019. Similarly, the Belfer Center's Cybersecurity Campaign Handbook (https://www.iri.org/sites/default/files/european_campaign_playbook_-_web.pdf), developed in partnership with NDI and IRI, provides context and clear guidance for campaigns facing a variety of cybersecurity issues, including disinformation and hacking. In terms of more concrete activities, DRG practitioners are building media monitoring into existing programs, including election observation (<https://www.power3point0.org/2018/03/06/fighting-fiction-countering-disinformation-through-election-monitoring/>). Grafting media monitoring onto existing program models and activities is a promising approach that could allow DRG programs to counter disinformation at scale. However, a potential drawback of this approach is that it focuses intervention on election cycles, while both the content and tactics transcend election cycles and operate over long periods of time.¹⁶ With this in mind, program designers and funders should consider support for efforts that bridge elections, and often, go further than the life of a standard DRG program.

Ultimately the real-world effects of content awareness and fact-checking programs are unclear. Academic research suggests that while fact-checking can change individual attitudes under very specific circumstances, it also has the potential to cause blowback or retrenchment – increased belief in the material that was fact-checked in the first place.¹⁷ Furthermore, there appears to be

relatively little research on whether fact-checking deters the proliferation of disinformation among political elites. Anecdotally, fact-checking may lead politicians to attempt to discredit the source, rather than change their behavior.¹⁸ Ultimately, an accounting of any deterrent effect of fact-checking program approaches will require donors and implementers to evaluate the impact of these programs more rigorously.

In any case, the existence of factchecking, verification outlets, or awareness building alone is likely not sufficient to change political actors' behavior regarding false statements or disinformation. In Ukraine, for example, research suggests that audiences for prominent fact-checking outlets were constrained geographically. The primary audiences tended to be younger, more urban, internet-connected, educated, and wealthy, and already inclined to monitor and sanction disinformation on their own.²⁰ Fact-checking and verification programs should therefore pay close attention to deliberately expanding audiences to include populations that might otherwise lack the opportunity or resources to access high quality information. These programs should also consider efforts to make elected officials themselves conscious of their monitoring mechanisms and audience reach. If candidates or elected officials are confident that the products of these outlets are not accessible to, or used by, their specific constituencies, these programs will be less effective in serving an accountability function.

RESEARCH PROGRAMS ON



HIGHLIGHT

CEPPS research identified dozens of programs that support fact-checking outlets across countries. For specific examples, consult the program repository and the [Poynter Institute International Fact-Checking Network](https://www.poynter.org/ifcn/) (<https://www.poynter.org/ifcn/>).

DISINFORMATION VULNERABILITY AND RESILIENCE

These programs focus on targets of disinformation, examining aspects of their background, the kinds of disinformation they respond to and other demographic factors to understand how they are susceptible to or can resist false content. For research programs with political party partners, these programs generally operate from a theory of change that hypothesizes that if there is a

greater awareness of organization vulnerabilities to disinformation, then political party officials will be motivated to improve their party's resilience. Two prominent examples of DRG programs that aim to provide research on vulnerability and resilience to policymakers, including elected officials and political parties, are IRI's Beacon Project, and NDI's INFO/tegrity Initiative.

IRI's Beacon Project (<https://www.iribeaconproject.org/>) supports original research into disinformation vulnerability and resilience with public opinion research, analytical pieces, narrative monitoring, and mainstream and social media monitoring through in-house expertise and in collaboration with local partners in Europe. These research products are shared among broad coalitions of stakeholders and applied in programmatic responses to disinformation narratives and through engagement with policymakers at the local, national, and European Union levels. Similarly, NDI's INFO/tegrity Initiative (<https://www.ndi.org/infotegrity>) commissions original research on vulnerabilities to disinformation, which in turn strengthens programming to build resilience, in partnership with political parties, social media platforms, and technology firms. Finally, DRG practitioners are increasingly working with academic partners to produce research on disinformation to improve programmatic approaches to improving resilience. For example, the Defending Digital Democracy (<https://www.belfercenter.org/D3P#!about>) project at Harvard University's Belfer Center connects academic research on disinformation threats and vulnerability to governments, CSOs, technology firms, and political organizations.

PROGRAMS FOR UNDERSTANDING THE SPREAD OF DISINFORMATION ONLINE

Researchers and programmers look to understand the roots of disinformation campaigns online by studying datasets of social media content to understand the virality of certain kinds of content, communities, and individual users' roles.

Disinformation is a cheap, effective, campaign tactic that usually goes entirely undetected, making the reputational cost for political party use of disinformation effectively nil. Several programs have recently emerged that track the use of content farms, false amplification, buying of followers/likes, troll armies, and other tactics by political actors. This programmatic approach has been supported by the growing accessibility of digital forensics research skills; increasing awareness among local actors of the role disinformation can play in political campaigns; and, due to concerns about malign foreign disinformation during elections, the investment made in content archiving technologies, social media mapping and graphing, and media monitoring platforms. This approach focuses on the behavior component of disinformation – it does not attempt to assess the veracity of the content being produced or amplified.

The implicit theory of change behind this work is that exposing the use of disinformation by political parties during campaigns will have some reputational cost, reducing their ability to deploy

disinformation tactics with impunity and damaging the electoral prospects of those who do.

Given that this approach is content agnostic, it is the one that most lends itself to changes in election rules. By exposing the tactics that political campaigns use that are most harmful to democratic integrity, election management bodies can explicitly forbid the use of those tactics during an election period.

PROGRAMS COMBATING HATE SPEECH, INCITEMENT, AND POLARIZATION

A component of disinformation and information integrity is the use of hate speech, often in combination with false information, to incite, suppress, or polarize users. This kind of program often exists separately from others focused on disinformation but could be evaluated as another potential response.

Hate speech, stereotyping, rumors, trolling, online harassment, and doxing are mechanisms through which parties might perform their *mobilization* function. Particularly in environments with pronounced political, social, or economic cleavages, the propagation of inflammatory information may serve to activate supporters or demobilize supporters of opposition parties. Both with respect to domestic and foreign campaigns, disinformation in this vein attempts to exacerbate these existing cleavages. Marginalized groups, including (but not limited to) women, ethnic, religious, or linguistic minorities, and LGBTI citizens are common targets of these campaigns, particularly where the perpetrators aim to scapegoat vulnerable groups for policy failures, or where perpetrators aim to deter participation of these groups in the political process, either by candidacy or voting. Indeed, across contexts, online violence against women, including hate speech and threats, infliction of embarrassment and reputational risks, and sexualized distortion, constituted a significant barrier to women's participation in the political process by causing silence, self-censorship, and withdrawal from political engagement, both for the immediate targets, and by deterring women's participation generally.²¹

Furthermore, these tactics can also help mobilize supporters by drawing on fear or anxiety around changing social hierarchies. Importantly, political communications framed as stereotypes can increase acceptance of false information about the group being stereotyped.²² This appeal of stereotypes creates a powerful incentive for political parties and politicians to attack vulnerable groups with disinformation in ways that are not experienced by members of favored in-groups.

Programmatically, DRG programs can counteract these effects by first acknowledging that disinformation disproportionately harms groups that have been historically marginalized in specific contexts, and by encouraging political party partners to engage in messaging that might improve supporters' attitudes toward vulnerable groups.²³ For example, the Westminster Foundation for Democracy Uganda (<https://www.wfd.org/2020/08/05/empowering-women->

[candidates-to-run-successful-campaigns-ahead-of-the-2021-general-elections-in-uganda/](#)) office organized an e-conference for over 150 women candidates for elected office in Uganda, with a focus on navigating social barriers to political participation, including misinformation and cyberbullying. Similarly, the [Women's Democracy Network \(https://www.wdn.org/\)](https://www.wdn.org/) is a global network of chapters that share best practices on identifying and overcoming barriers to women's political participation. NDI has several programs geared toward identifying and overcoming social barriers to participation within political parties specifically, including the [Win with Women initiative \(https://www.ndi.org/win-with-women-building-inclusive-21st-century-parties\)](https://www.ndi.org/win-with-women-building-inclusive-21st-century-parties) and the [#NotTheCost campaign \(https://www.ndi.org/not-the-cost\)](https://www.ndi.org/not-the-cost), designed to mitigate discrimination, harassment, violence, and other forms of backlash against women's political participation. Similarly, [NDI's safety planning tool \(https://think10.demcloud.org/\)](https://think10.demcloud.org/) provides a mechanism through which women who participate in politics can privately assess their security and make a plan to increase their safety, especially with respect to harassment, public shaming, threats and abuse, physical and sexual assault, economic violence, and pressure to leave politics, both in online and offline spaces.



KEY RESOURCE

Network Approaches to Scaling Best Practices

[Poynter Foundation International Fact-Checking Network](https://www.poynter.org/about-the-international-fact-checking-network/)

[\(https://www.poynter.org/about-the-international-fact-checking-network/\)](https://www.poynter.org/about-the-international-fact-checking-network/): The International Fact-Checking Network (IFCN) is a forum for fact-checkers worldwide hosted by the Poynter Institute for Media Studies. These organizations fact-check statements by public figures, major institutions, and other widely circulated claims of interest to society. The IFCN Model is further explored in [the norms and standards section \(/topics/norms/5-codes-conduct-researchers-fact-checkers-journalists-media-monitors-and-others#IFCNnormsstandards\)](/topics/norms/5-codes-conduct-researchers-fact-checkers-journalists-media-monitors-and-others#IFCNnormsstandards) and :

POLICY

- Monitors trends and formats in fact-checking worldwide, publishing regular articles on the dedicated [Poynter.org channel](https://www.poynter.org/channel) (<https://www.poynter.org/category/fact-checking/>).
- Provides training [resources](http://www.newsu.org/courses/fact-checking) (<http://www.newsu.org/courses/fact-checking>) for fact-checkers.
- Supports collaborative efforts in international fact-checking, including fellowships.
- Convenes a yearly conference ([Global Fact](https://www.washingtonpost.com/news/fact-checker/wp/2017/07/14/fighting-falsehoods-around-the-world-a-dispatch-on-the-global-fact-checking-movement/) (<https://www.washingtonpost.com/news/fact-checker/wp/2017/07/14/fighting-falsehoods-around-the-world-a-dispatch-on-the-global-fact-checking-movement/>)).
- Is the home of the fact-checkers' [code of principles](https://www.poynter.org/fact-checkers-code-of-principles/) (<https://www.poynter.org/fact-checkers-code-of-principles/>).

RECOMMENDATIONS AND REFORM/ SHARING AND SCALING GOOD PRACTICE IN PROGRAMMATIC RESPONSES

Programs that address policies around online systems, social media, and the internet can help define new rules that can reduce the impact of disinformation. A key role for DRG donors and implementing partners is to use their convening power to connect diverse stakeholders to share lessons learned and best practices, within and across countries and programs. It is important to note that many of the programs discussed above also have an important convening function – they are often deliberately designed to share best practices between key stakeholders, including politicians and political organizations, elected officials, civil servants, CSOs, media outlets, and

technology firms. These convening activities hypothetically could improve outcomes through two mechanisms. First, the exchange of lessons learned and best practices could increase the skills, knowledge, capacity, or willingness of political party partners to refrain from the use of disinformation, or to take steps to improve party resilience. Second, these convening activities could serve an important coordination function. Recall that one important implication of thinking of disinformation as a tragedy of the commons is that political parties and candidates might be willing to forgo the political advantages of disinformation if they could be confident their political opponents would do the same. Programs that provide regular, scheduled, ongoing interaction between political opponents could increase confidence that political competitors are not cheating.

IFES's Regional Elections Administration and Political Process Strengthening (REAPPS I and II) (<https://www.ifes.org/REAPPS>) programs in Central and Eastern Europe provide an example of how DRG support programs can facilitate this kind of coordination over a relatively long period of time. The program's thematic focus on information security and explicit attention to cross-sectoral and cross-border networking addresses both technical approaches and underlying political incentives.

Design 4 Democracy Coalition (<https://d4dcoalition.org/index.php/our-story>): The D4D Coalition aims to ensure that information technology and social media play a proactive role in supporting democracy and human rights globally. The coalition partners create programs and trainings and coordinate between members to promote the safe and responsible use of technology to advance open, democratic politics and accountable, transparent governments.

HELPING PARTIES PROTECT THE INTEGRITY OF POLITICAL INFORMATION

3. POLICY RECOMMENDATIONS (/TOPICS/PARTIES/3-POLICY-RECOMMENDATIONS)

POLICY RECOMMENDATIONS

- When implementing these programmatic approaches, consider political incentives in addition to technical solutions.

- Consider an inclusive, gender-sensitive landscape analysis or a political economy analysis to identify how the structure of social cleavages creates incentives and opportunities for candidates or political parties to exploit context-specific norms and stereotypes around gender identity, ethnic or religious identities, sexual orientation, and groups that have been historically marginalized within that context.
- Programmatic interventions should account for diverging interests within parties – parties are composed of functionaries, elected officials, interest groups, formal members, supporters, and voters – each of which may have unique incentives to propagate or take advantage of disinformation.
- The collective action problem of disinformation makes one-off interactions with single partners difficult – consider implementing technical programs with regular, ongoing interaction between all relevant parties to increase confidence that competitors are not “cheating.”
- Relatedly, use convening power of donors or implementing organizations to bring relevant actors to the table.
- Consider pacts or pledges, especially in pre-election periods, in which all major parties commit to mitigating disinformation. Importantly, the agreement itself is cheap talk, but pay careful attention to design of institutions, both within the pact and externally, to monitoring compliance.
- There is limited evidence for effectiveness of common counter-disinformation program approaches with a focus on political parties and political competition, including media literacy, fact-checking, and content labeling. That there is limited evidence does not necessarily imply these programs do not work, only that DRG funders and implementing partners should invest in the rigorous evaluation of these programs to determine their impact on key outcomes like political knowledge, attitudes and beliefs, polarization, propensity to engage in hate speech or harassment, and political behavior like voting, and to identify what design elements distinguish effective programs from ineffective ones.
- DRG program responses have tended to lag political parties’ use of sophisticated technologies like data harvesting, microtargeting, deep fakes and AI generated content. Funders and implementing partners should consider the use of innovation funds to generate concepts for responses to mitigate the potentially harmful effects of these tools, and to rigorously evaluate impact.

PLATFORM SPECIFIC ENGAGEMENT FOR INFORMATION INTEGRITY

0. OVERVIEW - PLATFORMS (/TOPICS/PLATFORMS/0-OVERVIEW- PLATFORMS)

Written by Vera Zakem, Senior Technology and Policy Advisor at the Institute for Security and Technology and CEO Zakem Global Strategies, Kip Wainscott, Senior Advisor for the National Democratic Institute, and Daniel Arnaudo, Advisor for Information Strategies at the National Democratic Institute

Digital platforms have become prominent resources for sharing political information, organizing communities, and communicating on matters of public concern. However, these platforms have undertaken a mix of responses and approaches to counter the growing prevalence of disinformation and misinformation affecting the information ecosystem. With a broad spectrum of communities struggling to mitigate the harmful effects of disinformation, hate speech coordinated influence operations, and related forms of harmful content, the private sector's access to privileged and proprietary data and metadata often uniquely positions them to understand these challenges.

A number of prominent social media companies and messaging platforms are leveraging their abundant data to help inform responses to disinformation and misinformation campaigns. These responses vary widely in character and efficacy, but can generally be characterized as falling into one of the following three categories:

1. [policies, product interventions, enforcement measures to limit the spread of disinformation](/topics/platforms/1-interventions-and-responses-limit-or-curtail-disinformation-and-misinformation) (</topics/platforms/1-interventions-and-responses-limit-or-curtail-disinformation-and-misinformation>);
2. [policies and product features to provide users with greater access to authoritative information, data, or context](/topics/platforms/2-strategies-boost-access-authoritative-information-and-data) (</topics/platforms/2-strategies-boost-access-authoritative-information-and-data>); and
3. [efforts to promote a stronger community response and societal resilience, including digital literacy and internet access, to disinformation and misinformation](/topics/platforms/3-efforts-promote-resiliency-digital-literacy-and-stronger-community-responses) (</topics/platforms/3-efforts-promote-resiliency-digital-literacy-and-stronger-community-responses>).

Many platforms have implemented new policies or changes in the enforcement of previously implemented policies in response to disinformation related to the COVID-19 pandemic, the 2020 U.S. Presidential Election, and the January 6th assault on the U.S. Capitol. With an increase of false information [related to COVID-19 \(https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation\)](https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation), fact-checking increased 900 percent from January to March 2020, according to an [Oxford University study \(https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation\)](https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation). The World Health Organization has characterized this spread of misinformation about COVID-19 as an “[infodemic \(https://time.com/5822372/facebook-coronavirus-misinformation/\)](https://time.com/5822372/facebook-coronavirus-misinformation/),” which [incidentally occurred \(https://pen.org/the-first-wave-social-media-platforms-responding-to-covid-19/\)](https://pen.org/the-first-wave-social-media-platforms-responding-to-covid-19/) at a time of increased social media use as many people have been restricted to their homes during the pandemic.

In addition, the 2020 U.S. presidential election inspired major platforms to [update their policies \(https://foundation.mozilla.org/en/campaigns/election-policy-tracker/\)](https://foundation.mozilla.org/en/campaigns/election-policy-tracker/) related to coordinated inauthentic behavior, manipulated media, and disinformation campaigns targeting voters and candidates. Similarly, the January 6th attack on the U.S. Capitol has motivated social media platforms to once again [reexamine and update \(https://abcnews.go.com/Business/wireStory/social-media-crackdown-continues-siege-us-capitol-75198967\)](https://abcnews.go.com/Business/wireStory/social-media-crackdown-continues-siege-us-capitol-75198967) their policies, and their enforcement, as it relates to disinformation and the potential risks of offline violence.

This chapter examines platform responses in greater detail in order to provide a foundational understanding of steps social media platforms and encryption services use to address disinformation. Across all of these various approaches, it is important to note that social media policies and enforcement actions are constantly evolving as the threat landscape constantly changes. To better understand these changes, most prominent platforms, including Twitter, YouTube, and Facebook, publish regularly transparency reports that provide data to users, policymakers, peers, and civil society stakeholders about how these platforms update their policies, their enforcement strategies, and product features to respond to the dynamic threat landscape and societal challenges. To help account for these ever-moving dynamics, and as highlighted throughout this chapter, platforms have in many cases partnered with local groups, civil society organizations, media, academics, and other researchers to design responses to these challenges in the online space. *Given the evolving nature of the threat landscape, the relevant policy, product, and enforcement actions, based on the information available as of the publication of this guide, are documented here.*

PLATFORM SPECIFIC ENGAGEMENT FOR INFORMATION

INTEGRITY

1. INTERVENTIONS AND RESPONSES TO LIMIT OR CURTAIL DISINFORMATION AND MISINFORMATION (/TOPICS/PLATFORMS/1-INTERVENTIONS-AND-RESPONSES-LIMIT-OR-CURTAIL-DISINFORMATION-AND-MISINFORMATION)

In recognition of disinformation's prevalence and potential to cause harm, many social media platforms have taken actions to limit, remove, or combat both disinformation and misinformation in various ways. These responses are typically informed by platforms' policies regarding content and behavior and mostly operationalized by product features and technical or human intervention. This section examines the variety of approaches companies have taken to address digital disinformation on their platforms.

Sometimes the moderation of content on social media comes under the pretense of combating disinformation when actually it is in service of illiberal government objectives (<https://freedomhouse.org/article/rise-digital-authoritarianism-fake-news-data-collection-and-challenge-democracy>). It is important to note that platforms controlled by companies in authoritarian countries often remove disinformation and other harmful content. This can raise important censorship concerns particularly when the harm is defined as criticism directed toward the government under which the company operates.

(A). Platform Policies on Disinformation and Misinformation

A handful of the world's largest and most popular social media companies have developed policies and community standards (<https://carnegieendowment.org/2021/04/01/how-social-media-platforms-community-standards-address-influence-operations-pub-84201>) to address disinformation and misinformation. This section examines some of the most significant private sector policy responses to disinformation, including from Facebook, Twitter, and YouTube, as well as growing companies such as Tik Tok and others.

1. Facebook Policies

At Facebook, user activities are governed by a set of policies known as Community Standards (<https://www.facebook.com/communitystandards/>). This set of rules does not presently ban disinformation or misinformation in general terms; however, it does feature several prohibitions that may apply to countering disinformation and misinformation in specific contexts. For example,

the Community Standards prohibit content (https://www.facebook.com/communitystandards/coordinating_harm_publicizing_crime) that misrepresents information about voting or elections, incites violence (https://www.facebook.com/communitystandards/credible_violence), promotes hate speech (https://www.facebook.com/communitystandards/hate_speech), or includes misinformation related to the Covid-19 pandemic (<https://about.fb.com/news/2020/03/combating-covid-19-misinformation/>). Also, the Community Standards prohibit "Coordinated Inauthentic Behavior," (https://www.facebook.com/communitystandards/inauthentic_behavior/) which is defined to generally prohibit activities that are characteristic of large-scale information operations on the platform. Once detected, networks participating in coordinated inauthentic behavior are removed. Also, Facebook has begun to develop policies (<https://about.fb.com/news/2019/10/inside-feed-womens-safety/>), engaging with experts (<https://about.fb.com/news/2021/06/partnering-with-experts-to-promote-womens-safety/>), and developing technology (<https://www.facebook.com/safety/tools>) to increase the safety of women on its platform and their family of apps. Rules against harassment, unwanted messaging, and non-consensual intimate imagery that is disproportionately targeted towards women are all part of Facebook's efforts to make women feel safer. However more work still needs to be done, as the burden often falls on women to report abuse and manage their safety on the platform.

Outside of these specific contexts, Facebook's Community Standards include an acknowledgment that while disinformation is not inherently prohibited, the company has a responsibility to reduce the spread of "false news." (https://www.facebook.com/communitystandards/false_news/) In operationalizing this responsibility, Facebook commits to algorithmically reduce (or down-rank) the distribution of such content, in addition to other steps to mitigate its impact and disincentivize its spread. The company has also developed a policy of removing particular categories of manipulated media (https://www.facebook.com/communitystandards/manipulated_media) that may mislead users; however, the policy is limited in scope. It extends only to media that is the product of artificial intelligence or machine learning and includes an allowance for any media deemed to be satire or content that edits, omits, or changes the order of words that were actually said.

It is worth recognizing that while Facebook's policies generally apply to all users, the company notes that "[i]n some cases, we allow content which would otherwise go against our Community Standards – if it is newsworthy and in the public interest" (<https://www.facebook.com/communitystandards/introduction>). The company has further indicated that speech by politicians (<https://about.fb.com/news/2019/09/elections-and-political-speech/>) will generally be treated as within the scope of this newsworthiness exception, and therefore not subject to removal. Such posts are, however, subject to labeling (<https://www.forbes.com/sites/rachelsandler/2020/06/26/in-reversal-zuckerberg-says-facebook-will-label-newsworthy-posts-that-violate-its-rules/?sh=3e687d3e7340>) that indicates that the posts violate the Community Standards. In recent years, Facebook has taken steps to remove political speech and deplatform politicians, including former President Donald Trump in the wake of the January 6th attack on the U.S. Capitol. Following the attacks, Facebook made the decision to remove President Donald Trump's account from the platform indefinitely. The Oversight Board upheld the

decision, but criticized the open ended nature (<https://transparency.fb.com/oversight/oversight-board-cases/former-president-trump-suspension-from-facebook/>) of the suspension, so Facebook limited the suspension to two years (<https://www.npr.org/2021/06/04/1003284948/trump-suspended-from-facebook-for-2-years>). In 2018, Facebook also de-platformed Min Aung Hlaing (<https://about.fb.com/news/2018/08/removing-myanmar-officials/>) and senior Myanmar military leaders for conducting disinformation campaigns and inciting ethnic violence.

2. Twitter Policies

The Twitter Rules (<https://help.twitter.com/en/rules-and-policies/twitter-rules>) govern permissible content on Twitter, and while there is no general policy on misinformation, the Rules do include several provisions to address false or misleading content (<https://help.twitter.com/en/rules-and-policies/election-integrity-policy>) and behavior in specific contexts. Twitter's policies prohibit disinformation and other content that may suppress participation or mislead people about when, where, or how to participate in a civic process (<https://help.twitter.com/en/rules-and-policies/election-integrity-policy>); content that includes hate speech (<https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy>) or incites violence or harassment; or content that goes directly against guidance from authoritative sources (https://blog.twitter.com/en_us/topics/product/2020/updates-our-approach-to-misleading-information.html) of global and local public health information. Twitter also prohibits inauthentic behavior and spam (<https://help.twitter.com/en/rules-and-policies/platform-manipulation>), which is an element of information operations that makes use of disinformation and other forms of manipulative content. Related to disinformation, Twitter has updated its hateful conduct policy to prohibit language that dehumanizes people on the basis of race, ethnicity, and national origin.

Following public consultation (https://blog.twitter.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media.html), Twitter has also adopted a policy regarding sharing synthetic or manipulated media (<https://help.twitter.com/en/rules-and-policies/manipulated-media>) that may mislead users. The policy requires an evaluation of three elements, including whether (1) the media itself is manipulated (or synthetic); (2) the media is being shared in a deceptive or misleading manner; and (3) the content risks causing serious harm (including users' physical safety, the risk of mass violence or civil unrest, and any threats to the privacy or ability of a person or group to freely express themselves or participate in civic events). If all three elements of the policy are met, including a determination that the content is likely to cause serious harm, Twitter will remove the content. If only some of the elements are met, Twitter may label the manipulated content, (https://blog.twitter.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media.html) warn users who try to share it or attach links to trusted fact-checking content to provide additional context.

In the context of electoral and political disinformation, Twitter's policies on elections (<https://help.twitter.com/en/rules-and-policies/election-integrity-policy>) explicitly prohibit misleading information about the voting process. Its rules note "You may not use Twitter's services for the purpose of manipulating or interfering in elections or other civic processes. This includes posting or sharing content that may suppress participation or mislead people about when, where,

or how to participate in a civic process." However, inaccurate statements about an elected or appointed official, candidate, or political party are excluded from this policy. Under these rules, Twitter has removed postings that feature disinformation about election processes, such as promoting the wrong voting day or false information about polling places--addressing content that EMBs election observers and others are increasingly working to monitor and report. It is notable that tweets from elected officials and politicians may be subject to Twitter's public interest intersectional.

Under their public interest exception, Twitter (https://blog.twitter.com/en_us/topics/company/2019/publicinterest.html) notes that it "may choose to leave up a Tweet from an elected or government official that would otherwise be taken down" and cites the public's interest in knowing about officials' actions and statements. When this exception applies, rather than remove the offending content, Twitter will "place it behind a notice providing context about the rule violation (<https://help.twitter.com/en/rules-and-policies/public-interest>) that allows people to click through to see the Tweet." The company has identified criteria (<https://help.twitter.com/en/rules-and-policies/public-interest>) for determining when a Tweet that violates its policies may be subject to this public interest exception, which includes:

1. The Tweet violated one or more of the platform's rules
2. Was posted by a verified account
3. The account has more than 100,000 followers
4. The account represents a current or potential member of a governmental or legislative body:
 - A. Current holders of an elected or appointed leadership position in a governmental or legislative body, OR
 - B. Candidates or nominees for political office, OR
 - C. Registered political parties

In considering how to apply this exception, the company has also developed and published a set of protocols (<https://help.twitter.com/en/rules-and-policies/public-interest>) for weighing the potential risk and severity of harm against the public-interest value of the Tweet. During the 2020 U.S. Presidential Election cycle, Twitter applied the public interest (https://blog.twitter.com/en_us/topics/company/2019/publicinterest.html) intersectional notice to many of former President Donald Trump's Tweets. Tweets that fall under this label display the following warning as shown below:



Donald J. Trump ✓
@realDonaldTrump



This Tweet violated the Twitter Rules about glorifying violence. However, Twitter has determined that it may be in the public's interest for the Tweet to remain accessible. [Learn more](#)

....These THUGS are dishonoring the memory of George Floyd, and I won't let that happen. Just spoke to Governor Tim Walz and told him that the Military is with him all the way. Any difficulty and we will assume control but, when the looting starts, the shooting starts. Thank you!

9:53 PM · May 28, 2020 · [Twitter for iPhone](#)

Following the January 6th attack on the U.S. Capitol and Tweets that showed incitement of violence, Twitter de-platformed former President Trump. The company published [a blog](https://blog.twitter.com/en_us/topics/company/2020/suspension.html) (https://blog.twitter.com/en_us/topics/company/2020/suspension.html) with its rationale.

Finally, it is also noteworthy that Twitter has demonstrated a willingness to develop policies in response to specific topics that present a risk of polluting the platform's information environment. For example, the company has implemented a special [set of policies](https://www.lawfareblog.com/twitter-brings-down-banhammer-qanon) (https://www.lawfareblog.com/twitter-brings-down-banhammer-qanon) for removing content related to the QAnon conspiracy theory and accounts that promulgate it. Since the start of the COVID-19 pandemic, Twitter has also developed policies to counter COVID-related disinformation and misinformation, including [COVID-19 misleading information policy](https://help.twitter.com/en/rules-and-policies/medical-misinformation-policy) (https://help.twitter.com/en/rules-and-policies/medical-misinformation-policy) that impacts health and public safety.

3. YouTube Policies

YouTube follows a three-strike policy that results in the suspension or termination of offending accounts related to disinformation. YouTube policies include several provisions relevant to disinformation in particular contexts, including content that aims to mislead voters about the time, place, means, or eligibility requirements for voting or participating in a [census](https://support.google.com/youtube/answer/2801973?hl=en) (https://support.google.com/youtube/answer/2801973?hl=en); that advances false claims related to the [eligibility requirements](https://support.google.com/youtube/answer/2801973?hl=en) (https://support.google.com/youtube/answer/2801973?hl=en) for political candidates to run for office and elected government officials to serve in office; or promotes violence or hatred against or harasses individuals or groups based on [intrinsic attributes](https://support.google.com/youtube/answer/2801939?hl=en&ref_topic=9282436) (https://support.google.com/youtube/answer/2801939?hl=en&ref_topic=9282436). In addition, YouTube has also expanded its anti-harassment policy that prohibits video creators from using hate speech and insults on the basis of gender, sexual orientation, or race.

Like other platforms, the rules also include a specific policy against disinformation regarding public health or [medical information \(https://support.google.com/youtube/answer/9891785?hl=en&ref_topic=9282436\)](https://support.google.com/youtube/answer/9891785?hl=en&ref_topic=9282436) in the context of the COVID-19 pandemic. As misleading YouTube videos about the coronavirus gained [62 million views \(https://www.bbc.com/news/technology-52662348\)](https://www.bbc.com/news/technology-52662348) in just the first few months of the pandemic, YouTube indicated it removed [“thousands and thousands \(https://www.bbc.com/news/technology-52662348\)”](https://www.bbc.com/news/technology-52662348) of videos spreading misinformation in violation of the platform’s policies. The platform reiterated its commitment to stopping the spread of such harmful content.

YouTube has also developed a policy regarding [manipulated media \(https://support.google.com/youtube/answer/2801973?hl=en\)](https://support.google.com/youtube/answer/2801973?hl=en), which prohibits content that has been technically manipulated or doctored in a way that misleads users (beyond clips taken out of context) and may pose a serious risk of egregious harm. To further mitigate risks of manipulation or disinformation campaigns, YouTube also has policies that prohibit [account impersonation \(https://support.google.com/youtube/answer/2801947?hl=en\)](https://support.google.com/youtube/answer/2801947?hl=en), misrepresenting one’s country of origin, or concealing association with a government actor. These policies also prohibit [artificially increasing engagement metrics \(https://support.google.com/youtube/answer/3399767?hl=en\)](https://support.google.com/youtube/answer/3399767?hl=en), either through the use of automatic systems or by serving up videos to unsuspecting viewers.

4. TikTok Policies

In January 2020, TikTok implemented the ability for users to flag content as misinformation by selecting their new [‘misleading information category \(https://newsroom.tiktok.com/en-us/building-to-support-integrity\)’](https://newsroom.tiktok.com/en-us/building-to-support-integrity). Owned by Chinese company ByteDance, TikTok has been overshadowed by [privacy concerns \(https://www.reuters.com/article/us-eu-tech-tiktok/tiktoks-mayer-pledges-fake-news-fight-in-call-with-eus-breton-eu-official-says-idUSKBN23G2XM\)](https://www.reuters.com/article/us-eu-tech-tiktok/tiktoks-mayer-pledges-fake-news-fight-in-call-with-eus-breton-eu-official-says-idUSKBN23G2XM), as Chinese regulation requires companies to comply with [government requests \(https://www.cnn.com/2020/08/04/tech/tiktok-trump-ban-bytedance/index.html\)](https://www.cnn.com/2020/08/04/tech/tiktok-trump-ban-bytedance/index.html) to hand over data. In April 2020, TikTok released a statement regarding the company’s [handling of personal information \(https://newsroom.tiktok.com/en-gb/security-approach\)](https://newsroom.tiktok.com/en-gb/security-approach), noting their “adherence to globally recognized security control standards like NIST CSF, ISO 27001 and SOC2,” goals towards more transparency, and limitations on the “number of employees who have access to user data.”

While such privacy concerns have loomed large in public debates involving the platform, disinformation is also a challenge that the company has been navigating. In response to these issues, TikTok has implemented [policies to prohibit misinformation \(https://newsroom.tiktok.com/en-us/combating-misinformation-and-election-interference-on-tiktok\)](https://newsroom.tiktok.com/en-us/combating-misinformation-and-election-interference-on-tiktok) that could cause harm to users, “including content that misleads people about elections or other civic processes, content distributed by disinformation campaigns, and health misinformation.” These [policies \(https://newsroom.tiktok.com/en-us/combating-misinformation-and-election-interference-on-tiktok\)](https://newsroom.tiktok.com/en-us/combating-misinformation-and-election-interference-on-tiktok) apply to all TikTok users (irrespective of whether they are public figures), and they are enforced through a combination of content removals, account bans, and making it more difficult to find harmful content, like misinformation and conspiracy theories, in the platform’s recommendations or search features. TikTok established a moderation policy “which prohibits synthetic or manipulated content that misleads users by [distorting the truth](#)

(<https://www.theverge.com/2020/8/5/21354829/tiktok-deepfakes-ban-misinformation-us-2020-election-interference>) of events in a way that could cause harm.” This includes banning deepfakes (<https://www.theverge.com/2020/8/5/21354829/tiktok-deepfakes-ban-misinformation-us-2020-election-interference>) in order to prevent the spread of disinformation.

5. Snapchat Policies

In January 2017, Snapchat created policies to combat the spread of disinformation for the first time. Snapchat implemented policies for its news providers on the platform’s Discover page in order to combat disinformation, as well as to regulate information that is considered inappropriate for minors. These new guidelines require news outlets to fact-check (<https://qz.com/892774/snapchat-quietly-updates-its-guidelines-to-prevent-fake-news-on-its-discover-platform/>) their articles before they can be displayed on the platform’s Discover page to prevent the spread of misleading information.

In an op-ed, Snapchat CEO Evan Spiegel described the platform as different (<https://www.axios.com/how-snapchat-is-separating-social-from-media-2513315946.html>) from other types of social media and many other platforms, saying “content designed to be shared by friends is not necessarily content designed to deliver accurate information.” The inherent difference between Snapchat and other platforms allows them to combat misinformation in a unique way. There isn’t a feed of information from users on Snapchat like there is with many other social media platforms—a distinction that makes Snapchat more comparable to a messaging app (<https://www.technologyreview.com/2017/11/29/147413/snapchat-has-a-plan-to-fight-fake-news-ripping-the-social-from-the-media/>). With Snapchat’s updates, the platform makes use of human editors (<https://www.poynter.org/fact-checking/2017/heres-why-snapchats-latest-update-further-insulates-it-from-fake-news/>) who monitor and regulate what is promoted on the Discover page, preventing the spread of false information.

In June 2020, Snapchat released a statement expressing solidarity with the Black community amid Black Lives Matter protests following the death of George Floyd. The platform (<https://www.washingtonpost.com/technology/2020/06/03/snapchat-stops-promoting-trump/>) said that it “may continue to allow divisive people to maintain an account on Snapchat, as long as the content that is published on Snapchat is consistent with our [Snapchat’s] community guidelines, but we [Snapchat] will not promote that account or content in any way.” Snapchat also announced that it would no longer promote (<https://www.nytimes.com/2020/06/03/technology/snapchat-trump.html>) President Trump’s tweets on its Discover home page, citing “that his public comments of the site could incite violence.”

6. VKontakte

While many of the largest platforms have adopted policies aimed at addressing disinformation, there are notable exceptions to this trend. For example, VKontakte is one of the most popular social media platforms in Russia (<https://www.loc.gov/law/help/social-media-disinformation/russia.php>) and has been cited for its use to spread disinformation, particularly in

Russian elections. The platform has also been cited for its use by Kremlin-backed groups to spread disinformation beyond Russia's borders, impacting other countries' elections, as in [Ukraine](https://www.theatlantic.com/international/archive/2019/04/russia-disinformation-ukraine-election/587179/) (<https://www.theatlantic.com/international/archive/2019/04/russia-disinformation-ukraine-election/587179/>). While the platform is frequently used as a means to spread disinformation, it does not appear that VKontakte is enforcing any policies to stop the spread of fake news.

7. Parler

Parler was created in 2018 and has often been dubbed as the “alternative” to Twitter for conservative voices, largely due to its focus on freedom of speech with only de minimis content moderation policies. This unrestricted, unmoderated speech has led to a rise in anti-Semitism, hate, disinformation and propaganda, and extremism. Parler has been linked to the coordinated planning of the January 6th [insurrection](https://www.vox.com/recode/2020/11/24/21579357/parler-app-trump-twitter-facebook-censorship) (<https://www.vox.com/recode/2020/11/24/21579357/parler-app-trump-twitter-facebook-censorship>) at the U.S. Capitol. In the aftermath of this event, multiple services including [Amazon](https://www.cnn.com/2021/01/09/tech/parler-suspended-apple-app-store/index.html), [Apple](https://www.cnn.com/2021/01/09/tech/parler-suspended-apple-app-store/index.html), and [Google](https://www.cnn.com/2021/01/09/tech/parler-suspended-apple-app-store/index.html) (<https://www.cnn.com/2021/01/09/tech/parler-suspended-apple-app-store/index.html>) booted Parler from their platforms due to the lack of content moderation and a serious risk of public safety. This move demonstrates the ways in which the wider marketplace can apply pressure on specific platforms to implement policies to combat disinformation and other harmful content. After discussions with [Amazon](https://www.bloomberg.com/news/articles/2021-02-15/parler-back-online-after-getting-boot-from-amazon-over-riot) (<https://www.bloomberg.com/news/articles/2021-02-15/parler-back-online-after-getting-boot-from-amazon-over-riot>) and [Apple](https://www.cnn.com/2021/04/19/tech/apple-parler-app-store/index.html) (<https://www.cnn.com/2021/04/19/tech/apple-parler-app-store/index.html>), Parler made changes to the app to better detect and monitor hate speech. [Google](https://www.androidheadlines.com/2021/04/google-parler-play-store.html) (<https://www.androidheadlines.com/2021/04/google-parler-play-store.html>) announced they would allow Parler to return if they made changes to the app consistent with Google's policies, but as of September 2021, Parler has still not been added back to the Google Play store.

(B). Technical and Product Interventions to Curtail Disinformation

Private sector platforms have developed a number of product features and technical interventions intended to help limit the spread of disinformation, while balancing the interests of free expression. The design and implementation of these mechanisms are highly dependent on the nature and functionality of specific platforms. This section examines responses across these platforms, including traditional social media services, image, and video sharing platforms, and messaging applications. Of note, one of the biggest issues these platforms have tried and continue to address across the board is virality - the speed at which information travels on these platforms. When virality is combined with algorithmic bias, it can lead to coordinated disinformation campaigns, civil unrest, and violent harm.

1. Traditional Social Media Services

Two of the world's largest social media companies, Facebook and Twitter, have implemented interventions and features that work either to suppress the virality of disinformation and alert users to its presence or create *friction* that impacts user behavior to slow the spread of false information within and across networks.

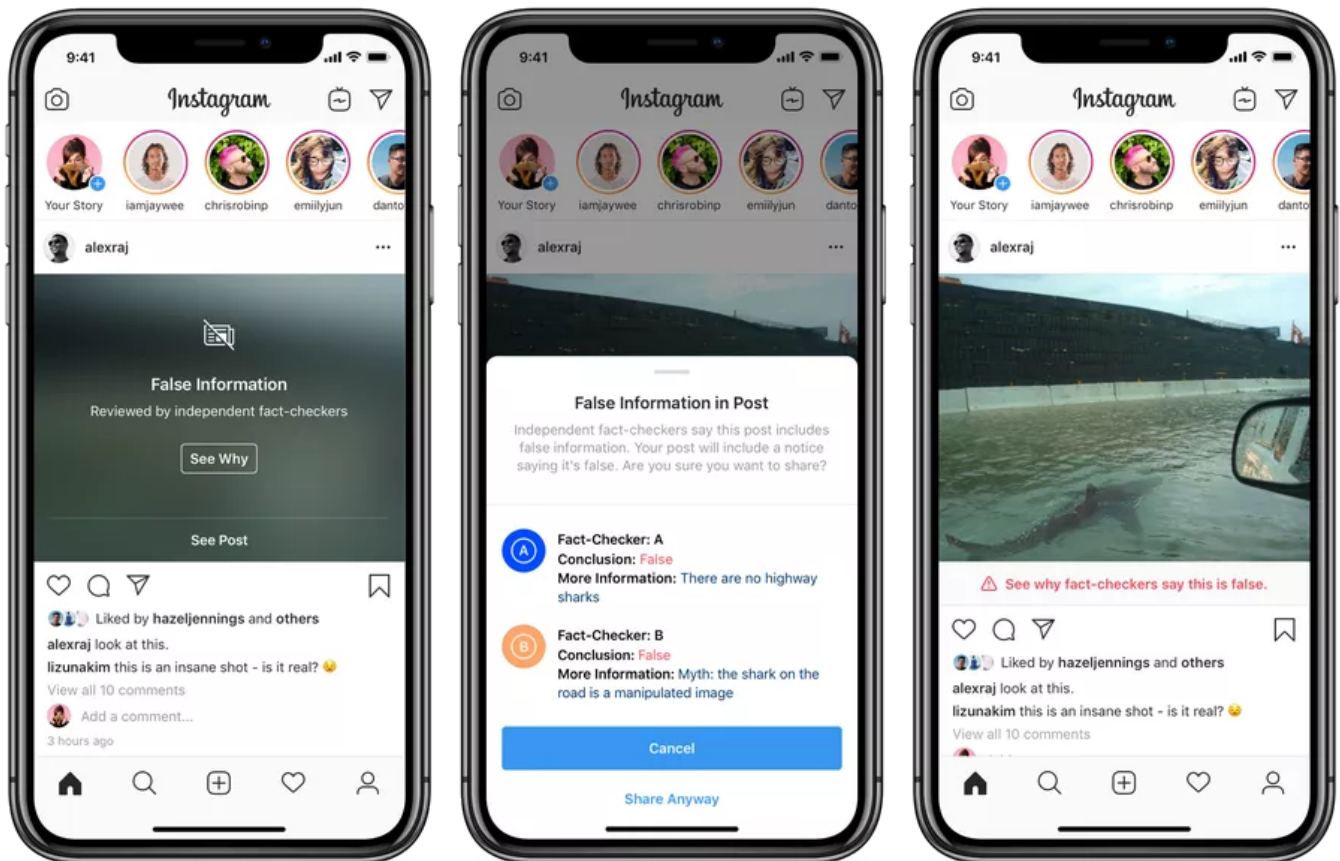
At Twitter, (https://www.buzzfeednews.com/article/janelytvynenko/new-twitter-article-feature-retweets?bftwnews&utm_term=4ldqpgc#4ldqpgc) product teams in 2020 began rolling out automated prompts that caution users against sharing links they have not themselves opened; this measure is intended to “promote informed discussion” and encourage users to evaluate information before sharing it. This follows the introduction of content labels and warnings, which the platform has affixed to Tweets that are not subject to removal under the platform’s policies (or under the company’s “public interest” exception, as described above) but which nonetheless may include misinformation or manipulated media (https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html). While these labels provide users with additional context (and are examined more thoroughly in the section of this chapter dedicated to such features (/topics/platforms/1-interventions-and-responses-limit-or-curtail-disinformation-and-misinformation#platformpolicies)), the labels themselves introduce a signal and potential friction that may impact a user’s decision to share or distribute misinformation.

Facebook’s technical efforts to curtail disinformation include using algorithmic strategies to “down-rank” false or disputed information, decreasing the content’s visibility in the News Feed and reducing the extent to which the post may be encountered organically. The company also applies these distribution limits (<https://about.fb.com/news/2018/05/hard-questions-false-news/>) against entire pages and websites that repeatedly share (<https://about.fb.com/news/2018/05/hard-questions-false-news/>) false news. The company has also begun to employ notifications to users who have engaged with certain misinformation (<https://about.fb.com/news/2018/05/hard-questions-false-news/>) and disinformation in limited contexts, such as in connection with a particular election or regarding health misinformation related to the COVID-19 pandemic. Although this intervention is in limited use, the company says these notifications (<https://about.fb.com/news/2018/05/hard-questions-false-news/>) are part of an effort to “help friends and family avoid false information.”

Both Twitter and Facebook utilize automation (https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html) for detecting certain types of misinformation and disinformation and enforcing content policies. These systems have played a more prevalent role during the pandemic as public health concerns have required human content moderators to disperse from offices. The companies similarly employ technical tools to assist in the detection of inauthentic activity on their platforms. While these efforts are not visible to users, the companies publicly disclose the fruits of these labors in periodic transparency reports which include data on account removals (https://blog.twitter.com/en_us/topics/company/2020/bot-or-not.html). These product features have been deployed globally, including in Sri Lanka, Myanmar, Nigeria, and other countries.

Platforms that share videos and images have also integrated features into their products to limit the spread of false information. On Instagram, a Facebook company, the platform removes content identified as misinformation from its Explore page and from hashtags; the platform also makes accounts that repeatedly post misinformation harder to find by filtering content from that

account from [searchable pages](https://about.fb.com/news/2019/10/update-on-election-integrity-efforts/). (<https://about.fb.com/news/2019/10/update-on-election-integrity-efforts/>). Examples of the ways Instagram has integrated curbing dis- and misinformation into their product development is shown below:



TikTok uses technology to augment its content moderation practices, particularly to assist in identifying inauthentic behavior, patterns, and accounts dedicated to spreading misleading or spam content. The company notes that its tools enforce their rules and make it more difficult to find harmful content, like misinformation and conspiracy theories, in the platform's recommendations or search features.

To support enforcement of its policies, [YouTube](https://www.youtube.com/howyoutubeworks/policies/community-guidelines/#detecting-violations) (<https://www.youtube.com/howyoutubeworks/policies/community-guidelines/#detecting-violations>) similarly employs technology, particularly machine learning, to augment its efforts. As the company notes among its policies, "machine learning is well-suited to detect patterns, which helps us to find content similar to other content we've already removed, even before it's viewed."

2. Messaging Applications

Messaging platforms have proven to be significant vectors for the proliferation of disinformation. The risks are particularly pronounced among closed, encrypted platforms, where companies are unable to monitor or review content.

Despite the complexity of the disinformation challenge on closed platforms, WhatsApp in particular has been developing technical approaches to mitigate the risks. Following a violent episode in India linked to viral messages on the platform being forwarded to large groups of up to 256 users at a time, [WhatsApp](https://www.theverge.com/2020/4/27/21238082/whatsapp-) (<https://www.theverge.com/2020/4/27/21238082/whatsapp->

[forward-message-limits-viral-misinformation-decline](#)) introduced limits on message forwarding in 2018 -- which prevent users from forwarding a message to more than five people -- as well as visual indicators to ensure that users can distinguish between forwarded messages and original content. More recently, in the context of the COVID-19 pandemic, WhatsApp further limited forwarding by announcing that [messages](#) (<https://www.theverge.com/2020/4/27/21238082/whatsapp-forward-message-limits-viral-misinformation-decline>) that have been forwarded more than five times can subsequently only be shared with one [chat](#) (<https://faq.whatsapp.com/general/chats/about-forwarding-limits/?lang=en>) at a time. While the encrypted nature of the platform makes it difficult to assess the impact of these restrictions on disinformation specifically, the company reports that the [limitations](#) (<https://www.theverge.com/2020/4/27/21238082/whatsapp-forward-message-limits-viral-misinformation-decline>) have reduced the spread of forwarded messages by 70%.

In addition to restricting forwarding behavior on the platform, WhatsApp has also developed systems for identifying and taking down automated accounts that send high volumes of messages. The platform is experimenting with methods to detect patterns in [messages](#) (https://www.cjr.org/tow_center/whatsapp-doesnt-have-to-break-encryption-to-beat-fake-news.php) through homomorphic encryption evaluation practices. These strategies may help to inform analysis and technical interventions related to disinformation campaigns in the future.

WhatsApp, owned by Facebook, is especially seeking to combat misinformation about COVID-19 as such content continues to go [viral](#) (<https://abcnews.go.com/Health/coronavirus-misinformation-whatsapp-viral-steps-combat-spread/story?id=69688321>). Efforts by the company have helped stop the [spread](#) (<https://www.forbes.com/sites/isabeltogoh/2020/04/27/whatsapp-viral-message-forwarding-drops-70-after-new-limits-to-stop-coronavirus-misinformation/#38a6a20d490d>) of COVID related dis and misinformation. WhatsApp has created a WHO Health Alert chatbot to provide accurate information about COVID-19. Users can text a phone number to access the chatbot. The chatbot provides basic information initially and allows users to ask questions on topics, including latest numbers, protection, mythbusters, travel advice and current news. This allows users to obtain accurate [information](#) (<https://www.protocol.com/coronavirus-instagram-tiktok-whatsapp-response>) and get direct answers to their questions. WhatsApp has provided [information](#) (<https://www.wired.com/story/whatsapp-coronavirus-who-information-app/>) through this service to over one million users.

3. Search Engines

[Google](#) (<https://fortune.com/2017/04/25/google-search-algorithm-fake-news/>) has implemented technical efforts to promote information integrity in search. Google changed its search algorithm to combat fake news dissemination and conspiracy theories. In a blog post, Google Vice President of Engineering Ben Gomes wrote that the company will “help surface more authoritative pages and demote low-quality content” in [searches](#) (<https://blog.google/products/search/our-latest-quality-improvements-search/>). In an effort to provide improved search guidelines, Google is adding real people to act as evaluators to “assess the quality of Google’s search results—give us

feedback on our [experiments \(https://blog.google/products/search/our-latest-quality-improvements-search/\)](https://blog.google/products/search/our-latest-quality-improvements-search/)." Google will also provide "direct feedback tools" to allow users to flag unhelpful, sensitive, or inappropriate content that appears in their searches.

PLATFORM SPECIFIC ENGAGEMENT FOR INFORMATION INTEGRITY

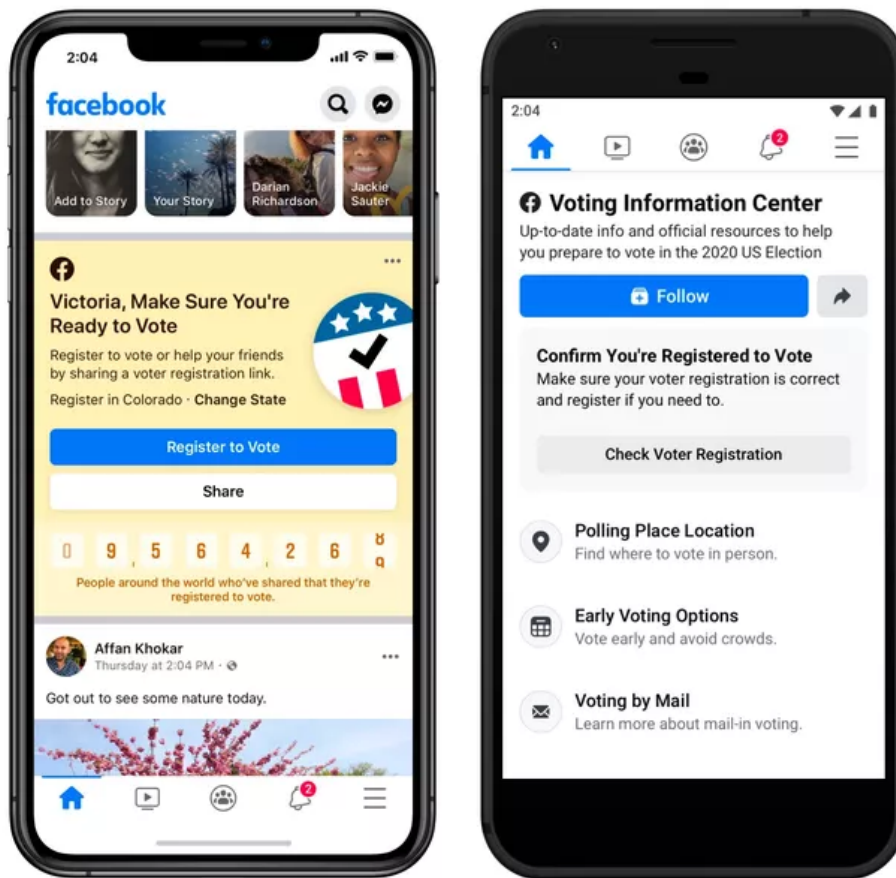
2. STRATEGIES TO BOOST ACCESS TO AUTHORITATIVE INFORMATION AND DATA (/TOPICS/PLATFORMS/2- STRATEGIES-BOOST-ACCESS- AUTHORITATIVE-INFORMATION-AND- DATA)

As private-sector technology platforms confront the issue of disinformation and misinformation across their services, one common strategy has been to provide users with greater access to authoritative information and contextualizing data. These strategies have, to date, included labeling content that may be misleading or harmful to users, directing users to official information sources on important topics like voting or public health, and providing researchers and civil society observers with access to tools and data to better understand the information environment across various digital services.

1. Facebook

Facebook has implemented a number of initiatives to improve access to data and authoritative information both for users and researchers. In the context of elections, for example, the company has introduced information labels that affix to any user content referencing "ballots" or "voting" (irrespective of the content's veracity). The labels direct users to official voting information and build on related efforts from different international contexts. For example, in Colombia during the election and peace process, Facebook created an Informed Voter button and Election Day reminders, which helped to spread credible information about the election process. In preparation for the 2019 local elections in Colombia, Facebook partnered with Colombia's National Electoral Council (CNE) to provide credible information about voting to citizens by including a voter button and including a reminder about voting. The *informed voter button*, as in other contexts, redirected the user to the local election authority for voter information about where and when they could vote

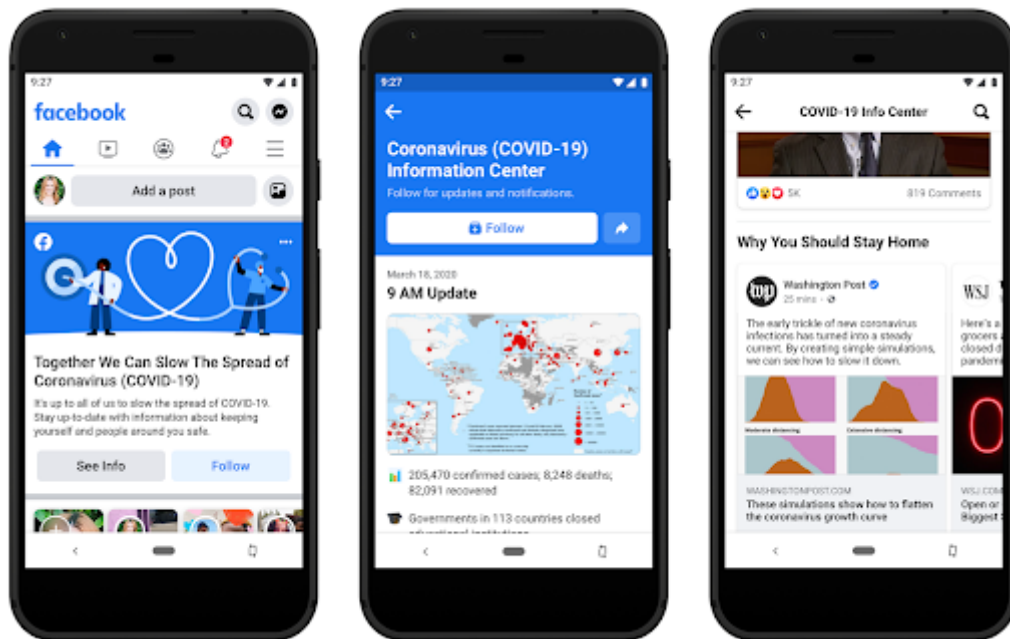
(<https://www.theguardian.com/technology/2018/apr/15/facebook-says-it-voter-button-is-good-for-turn-but-should-the-tech-giant-be-nudging-us-at-all>). Below, is an example of the voter information available on Facebook.



Facebook has also begun labeling certain state-controlled media to provide greater transparency (<https://about.fb.com/news/2020/06/labeling-state-controlled-media/>) on the sources of information on the platform. These labels currently appear on the platform's ad libraries and on Pages and will eventually be expanded to be more widely visible (<https://about.fb.com/news/2020/06/labeling-state-controlled-media/>). The labels build on transparency features already in operation on Facebook Pages, which include panels that provide context on how the page is being administered (including information about the users who manage the page and the countries from which they are operating), as well information about whether the page (<https://about.fb.com/news/2020/06/labeling-state-controlled-media/>) is state-controlled. It has expanded to include labels on state-sponsored media, a practice that was replicated by Twitter in August 2020 and also included political and media figures.

In response to the pandemic, in April 2020, Facebook announced that it will tell users if they have been exposed to misinformation (<https://www.politico.com/news/2020/04/16/facebook-fake-news-coronavirus-190054>) about COVID-19 and will direct users who have engaged with the misinformation to a website by the World Health Organization (<https://www.nbcnews.com/tech/tech-news/facebook-will-start-notifying-users-who-interact-coronavirus-misinformation-n1185146>) that debunks COVID-19 myths. Facebook is also removing COVID-19 misinformation (<https://about.facebook.com/actions/responding-to-covid-19/>) from the platform, based on guidance from public health authorities. Facebook created a Coronavirus Information Center (https://www.facebook.com/coronavirus_info/), which contains information

about the virus from public health officials and local leaders. Through these efforts, [Facebook](https://www.nbcnews.com/tech/tech-news/facebook-will-start-notifying-users-who-interact-coronavirus-misinformation-n1185146) (<https://www.nbcnews.com/tech/tech-news/facebook-will-start-notifying-users-who-interact-coronavirus-misinformation-n1185146>) and Instagram have directed over 2 billion people to reliable health information. The graphic below highlights Facebook's efforts to educate consumers about COVID-19 disinformation.



In support of research and analysis on the platform, Facebook enables greater access to data through its Crowdtangle application, which allows users to track pages and articles through a dashboard and downloadable datasets. [Crowdtangle](https://research.fb.com/blog/2020/07/crowdtangle-opens-public-application-for-academics/) (<https://research.fb.com/blog/2020/07/crowdtangle-opens-public-application-for-academics/>) is becoming available for academics and other researchers more openly. [Crowdtangle](https://research.fb.com/blog/2020/07/crowdtangle-opens-public-application-for-academics/) (<https://research.fb.com/blog/2020/07/crowdtangle-opens-public-application-for-academics/>) also has an open plug-in (<https://chrome.google.com/webstore/detail/crowdtangle-link-checker/klakndphagmmfklpelfkgjbjkimjihpmkh?hl=en>) for Chrome that allows users to understand the reach of articles throughout Facebook, Instagram, Reddit, and Twitter.

In addition, as advertising is a common vector for the spread of political and other forms of disinformation, Facebook has expanded access to advertising information through various ads databases and archives. This includes access to the [Ad Library API](https://techcrunch.com/2019/03/28/facebook-ads-library/) (<https://techcrunch.com/2019/03/28/facebook-ads-library/>) for researchers and those with Facebook developer accounts to study data about how ads are used in real time and to prevent misuse of the platform through targeted ads. The API allows researchers to access Facebook's dataset of content more directly through an automated system and creates a comprehensive mechanism for data collection and analysis.

2. WhatsApp

As an encrypted messaging platform, WhatsApp has only limited information available to users and researchers about activities on its services. However, WhatsApp has supported access to its API in order to support certain research initiatives. The company has expanded API access through the Zendesk system—particularly for groups connected to the First Draft Coalition, such

as Comprova in Brazil and CrossCheck in Nigeria. This approach has been utilized to collect data on political events, the spread of false information and hate speech, and other research goals. The International Fact-Checking Network¹ has also developed a collaboration with WhatsApp that includes access to the API for certain kinds of research, including an initiative launched in 2020 to [combat misinformation \(https://www.poynter.org/fact-checking/2020/ifcn-receives-1-million-from-whatsapp-to-support-fact-checkers-on-the-coronavirus-battlefront/\)](https://www.poynter.org/fact-checking/2020/ifcn-receives-1-million-from-whatsapp-to-support-fact-checkers-on-the-coronavirus-battlefront/), associated with the COVID-19 pandemic.

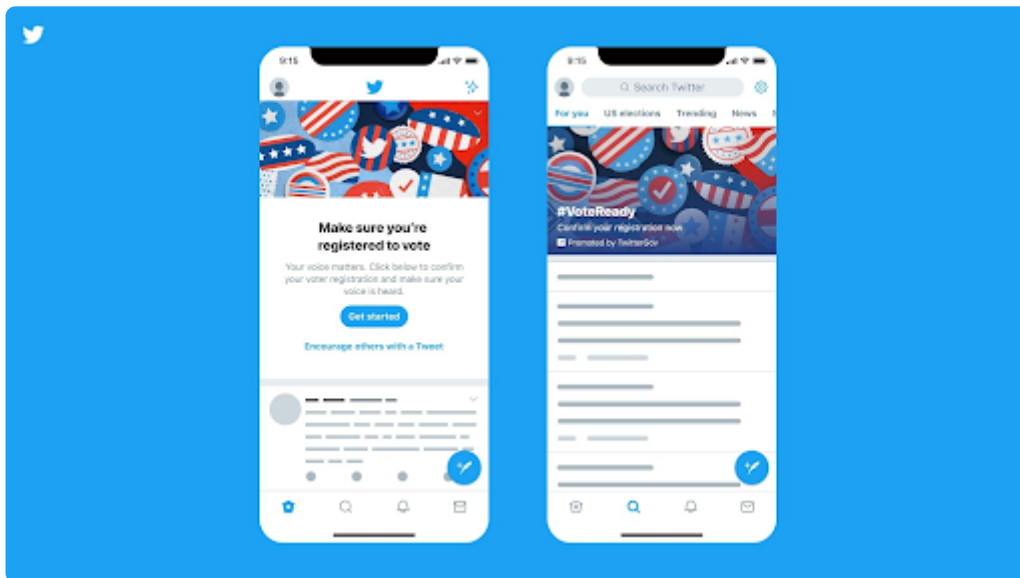
3. Twitter

Twitter has developed a number of policies, campaigns, and product features with the goal of providing users with access to credible and authoritative information. In 2020, Twitter undertook substantial efforts to provide users with access to credible information about elections, including the [U.S. Presidential Election \(https://blog.twitter.com/en_us/topics/company/2020/2020-election-changes.html\)](https://blog.twitter.com/en_us/topics/company/2020/2020-election-changes.html), so that users could reliably access accurate information on voting and the integrity of the election results. These efforts also included additional product features and enhancements to prevent users from sharing misleading information about voting.

Similarly, in connection with the COVID-19 pandemic, Twitter made [robust investments \(https://blog.twitter.com/en_us/topics/company/2020/covid-19.html#EmpoweringResearch\)](https://blog.twitter.com/en_us/topics/company/2020/covid-19.html#EmpoweringResearch) in ensuring users find reliable and credible public health information. For example, a [#KnowTheFacts search prompt \(https://blog.twitter.com/en_us/topics/company/2020/authoritative-information-about-novel-coronavirus.html\)](https://blog.twitter.com/en_us/topics/company/2020/authoritative-information-about-novel-coronavirus.html) has been translated in multiple languages and helps users find local, credible information and links to organizations that are working to curb COVID-19 threats. The company also updated its approach to address and contextualize misleading information about COVID-19 on its platform. For instance, Twitter announced in May 2020 that the company would [label misleading tweets \(https://www.cnn.com/2020/05/11/tech/twitter-coronavirus-misinformation/index.html\)](https://www.cnn.com/2020/05/11/tech/twitter-coronavirus-misinformation/index.html) about COVID-19 “to provide additional explanations or clarifications in situations where the risks of harm associated with a Tweet are less severe but where people may still be confused or misled by the content.” Twitter has also provided access to its API for researchers and academics to further study the public conversation surrounding COVID-19 in real time.

More broadly, Twitter has partnered [with UNESCO \(https://en.unesco.org/news/unesco-and-twitter-team-media-and-information-literacy\)](https://en.unesco.org/news/unesco-and-twitter-team-media-and-information-literacy) and the [OAS \(https://www.oas.org/en/sms/cicte/docs/20190913-DIGITAL-ENG-Alfabetismo-y-seguridad-digital-Twitter.pdf\)](https://www.oas.org/en/sms/cicte/docs/20190913-DIGITAL-ENG-Alfabetismo-y-seguridad-digital-Twitter.pdf) on guides to improve media literacy, as well as organizations around the world. Following its decision to [prohibit political advertising \(https://business.twitter.com/en/help/ads-policies/ads-content-policies/political-content.html\)](https://business.twitter.com/en/help/ads-policies/ads-content-policies/political-content.html), Twitter announced it was deprecating its [Ad Transparency Center \(https://business.twitter.com/en/help/ads-policies/product-policies/ads-transparency.html\)](https://business.twitter.com/en/help/ads-policies/product-policies/ads-transparency.html).

Below are examples of Twitter’s approach to providing credible information to users:



Know the facts

To make sure you get the best information on vaccinations, resources are available from the US Department of Health & Human Services.

Visit [vaccines.gov](https://www.vaccines.gov)

@HHSGov

Colombia's Election Management Body, the National Electoral Council (*Consejo Nacional Electoral* or CNE), has worked with Facebook and Twitter to actively promote access to credible information about the election, monitor disinformation, and provide enhanced product features informing, reminding, and educating voters about voting. The CNE has signed MOUs with both companies and worked actively to train its officials on monitoring online platforms and reporting content. CNE also worked in coordination with the companies during the 2019 local elections to train staff on the use of Twitter tools such as Tools, Periscope, Moments, Twitter Mirror, Q&A, Tweetdeck, and other best practices. CNE also helped these platforms set up an automated account to respond to elections queries, a hashtag, and enabled communications such as live videos throughout the election process. For CNE, partnering with Facebook and Twitter has been especially important, given that disinformation affects all people, including marginalized communities such as women, LGBTQ persons, and others.

4. Google

Google's "knowledge panels (<https://support.google.com/knowledgepanel/answer/9163198?hl=en>)" are boxes of information that appear when users search for people, places, things, and organizations that are in the "knowledge graph." These panels automatically generate boxes of information that provide a snapshot of information on a particular topic. While knowledge panels (<https://searchengineland.com/google-adds-new-knowledge-graph-learn-news-publishers-286394>) were created to provide information and address misinformation and fake news, they have been

the cause of [magnifying](https://www.theatlantic.com/technology/archive/2019/09/googles-knowledge-panels-are-magnifying-disinformation/598474/) some disinformation. Example of a knowledge panel is depicted below:

The image shows a Google search for "Eddie Aikau". The search bar at the top contains the name "Eddie Aikau" and shows "About 350,000 results (0.52 seconds)". Below the search bar, there are navigation tabs for "All", "News", "Videos", "Images", "Shopping", and "More". The main content area is divided into two columns. The left column features a Wikipedia snippet for "Eddie Aikau - Wikipedia", which includes his full name, birth and death dates, and a brief biography. Below this is a "People also ask" section with four questions: "How did Eddie Aikau die?", "What year did Eddie Aikau die?", "Is Eddie Aikau still alive?", and "What is the story of Eddie Aikau?". Underneath are three video thumbnails with titles like "How Eddie Aikau Became One of Surfing's Most Legendary ..." and "Who Was Eddie Aikau And What Happened To Him? Google ...". The right column is a knowledge panel for "Eddie Aikau" with the subtitle "Lifeguard". It contains a detailed biography, key facts such as "Born: May 4, 1946, Kahului, HI" and "Died: March 17, 1978, Hawaii", and a list of "People also search for" including Clyde Aikau, Duke Kahanamoku, Mark Foo, Nainoa Thompson, and Greg Noll.

5. YouTube

In order to provide users with accurate information, YouTube provides “[Breaking News](https://support.google.com/youtube/answer/9057101?hl=en)” and “[Top News](https://support.google.com/youtube/answer/9057101?hl=en)” features (<https://support.google.com/youtube/answer/9057101?hl=en>), which elevate information from verified news sources. As part of the company’s ongoing efforts, [YouTube](https://economictimes.indiatimes.com/magazines/panache/fighting-fake-news-youtube-to-show-information-panels-on-news-related-videos/articleshow/68302365.cms?from=mdr) (<https://economictimes.indiatimes.com/magazines/panache/fighting-fake-news-youtube-to-show-information-panels-on-news-related-videos/articleshow/68302365.cms?from=mdr>) has indicated that it is expanding the use of “information panels” to provide users with additional context from fact-checkers.

Youtube has also worked to label certain content during the COVID- 19 pandemic as questionable and has taken down content that was verifiably misleading, particularly related to the [Plandemic video](https://www.nytimes.com/2020/05/20/technology/plandemic-movie-youtube-facebook-coronavirus.html) (<https://www.nytimes.com/2020/05/20/technology/plandemic-movie-youtube-facebook-coronavirus.html>), which went viral in March 2020 as COVID-19 began spreading rapidly.

6. TikTok

To promote authoritative COVID-19 information in response to the spread of disinformation, TikTok has partnered with the World Health Organization (WHO) to create a landing page (<https://newsroom.tiktok.com/en-gb/taking-action-against-covid-19-vaccine-misinformation>) for the organization to provide accurate facts, safety information, Q&As, and informational videos about the pandemic. This partnership allows the WHO to provide information to a younger audience than many other social media platforms. TikTok's Head of Product said this partnership has allowed the platform to act "globally and comprehensively" to provide accurate information to its users. TikTok (<https://www.protocol.com/coronavirus-instagram-tiktok-whatsapp-response>) also revised its guidelines to denounce misleading information and flag inaccurate content.

7. Snapchat

To boost authoritative COVID-19 information, Snapchat implemented filters within its platform that feature verified information on reducing the risk of contracting COVID-19. While the platform allows independent creators to make filters, it will not allow misinformation to be included in them. Snapchat also announced the launch of a health and wellness initiative in response to user concern about COVID-19. The "Here for You" (<https://www.axios.com/exclusive-snapchat-expedites-wellness-push-in-response-to-virus-74aa5722-6401-4417-a1f8-df0f72931d14.html>) tool in the search bar will allow users to access information about mental health as well as information directly from the WHO, the CDC, the Crisis Text Line, the Ad Council, and the National Health Service.

PLATFORM SPECIFIC ENGAGEMENT FOR INFORMATION INTEGRITY

3. EFFORTS TO PROMOTE RESILIENCY, DIGITAL LITERACY, AND STRONGER COMMUNITY RESPONSES TO DISINFORMATION

(/TOPICS/PLATFORMS/3-EFFORTS-

PROMOTE-RESILIENCY-DIGITAL-LITERACY-AND-STRONGER-COMMUNITY-RESPONSES)

Collective action, community partnerships, and civil society engagement are important aspects of the private sector approach to addressing disinformation. These include individual companies' investments, engagement, and partnerships, as well as collaborative initiatives involving multiple companies. This section examines partnerships and initiatives undertaken by particular companies, as well as cross-sectoral and multi-stakeholder collaborations to combat disinformation.

A. COMPANY PARTNERSHIPS AND INITIATIVES

All major technology companies, such as Facebook, Google, and Twitter, have collaborated with civil society and others to combat disinformation, hate speech, and other harmful forms of content on their platforms. This section reviews some of the key initiatives they have undertaken to work with outside groups, particularly civil society organizations, on information space problems collectively.

1. Facebook

Facebook has developed a number of public-facing partnerships and initiatives aimed at supporting civil society and other stakeholders working to promote information integrity. Among its most notable announcements, Facebook has inaugurated an independent Oversight Board (<https://about.fb.com/news/2020/05/welcoming-the-oversight-board/>). The Board is composed of technology, human rights, and policy experts who have been given the authority to review difficult cases of speech that cause online harassment, hate, and spread disinformation and misinformation. As of the date of publishing this guidebook, the Oversight Board (<https://oversightboard.com/news/719406882003532-announcing-the-oversight-board-s-first-cases-and-appointment-of-trustees/>) has reviewed and made a determination on content moderation cases, including cases in China, Brazil, Malaysia, and the United States. This is significant, as the oversight board takes into account human rights, legal, and impact on society in reviewing difficult cases the platform may not be in the position to address.

The company has also invested in country-specific and regional initiatives. For example, WeThink Digital (<https://wethinkdigital.fb.com/partners/>) is a Facebook initiative to foster digital literacy through partnerships with civil society organizations, academia, and government agencies in various Asia-Pacific countries such as Indonesia, Myanmar, New Zealand, the Philippines, Sri Lanka, and Thailand. It includes public guides to user actions such as deactivating an account, digital learning modules, videos, and other pedagogical resources (<https://wethinkdigital.fb.com/resources/>). In the context of elections, in particular, Facebook has

also developed partnerships with election monitoring bodies, law enforcement, and other government institutions dedicated to the investigation of campaigns during electoral processes through the creation of a “war room (<https://www.nytimes.com/2018/09/19/technology/facebook-election-war-room.html>)” of dedicated staff in certain cases, such as the European Union (<https://www.politico.eu/article/facebook-european-election-war-room-dublin-political-advertising-misinformation-mark-zuckerberg/>), Ukraine (<https://www.rferl.org/a/leading-ukraine-candidate-zelensky-face-facebook-fakes-political-ad-rules/29828605.html>), Ireland, Singapore (<https://www.cnbc.com/2019/01/28/facebook-will-open-new-war-rooms-in-dublin-and-singapore.html>), Brazil (<https://www.cnet.com/news/facebook-sets-up-a-war-room-ahead-of-brazil-and-us-elections/>), and for the 2020 U.S. election (<https://www.cnbc.com/2018/11/26/facebook-wont-use-its-war-room-for-future-elections-report-says.html>), which they have since closed. According to the NDI case study on the role of social media platforms in enforcing policy decisions during elections, both Facebook and Twitter worked with the National Electoral Council (CNE) in Colombia during the electoral process.

In some countries, Facebook is partnering with third-party fact-checkers to review and rate the accuracy of articles and posts on the platform. As part of these efforts, in countries (<https://www.facebook.com/help/publisher/18222309230722>) such as Colombia, Indonesia, Ukraine, as well as various members of the EU and the United States, Facebook has commissioned (<https://newsroom.fb.com/news/2018/06/hard-questions-fact-checking/>) groups (<https://newsroom.fb.com/news/2018/06/hard-questions-fact-checking/>)—through what is described as “a thorough and rigorous application process” established by the IFCN²—to become trusted fact-checkers who vet content, provide input into the algorithms that define the news feed, and downgrade and flag content that is identified as false. In Colombia, for example—where partners include AFP Colombia, ColombiaCheck, and La Silla Vacía—a representative from one of these partners reflected on the value of working with Facebook and platform's more broadly: “I think the most important thing is to talk more closely with other platforms because the way to widen our reach is to work with them. Facebook has its problems but it reaches a lot of people and especially reaches the people that have shared false information, and if we could do something like that with Twitter, Instagram, or WhatsApp it would be great; that is the ideal next step for me.”³ Groups from more than 80 countries have partnered with Facebook in this way, underscoring the broad scope of this effort.

2. WhatsApp

While it is a closed platform, WhatsApp has supported researchers in developing studies of its platform as one of the principal means of community engagement. The studies include an interesting range of potential methodology and show how enhanced access can lead to interesting and important results for understanding the closed platform, especially



HIGHLIGHT

In Focus: Facebook’s Social Science One

how it is used in lesser-seen or known contexts. Many countries and regions are a black box, especially at a local level. Groups are closed, the platform is encrypted, and it is difficult to see and understand anything in terms of content moderation.

Abuse and online manipulation of WhatsApp through automated networks are common in many places. Local languages, dialects, and slang are not well known to moderators from different regions and countries. Violence against women online, in politics and elections, can have serious impacts on the political participation of the targeted individuals, as well as a chilling effect on the participation of women more broadly, and monitoring for hate speech should seek to understand methods of tracking local lexicons. The CEPPS partners have developed methodologies for tracking online hate speech against women and other marginalized groups, such as IFES's Violence Against Women in Elections (VAWIE) framework (<https://www.cepps.org/technical-leadership/gept-vawie-online-how-do-online-threats-of-violence-impact-womens-electoral-participation/>), or NDI's Votes without Violence Toolkit (<https://www.ndi.org/VAW-E>) and the Addressing Online Misogyny and Gendered Disinformation: A How-To Guide (<https://www.ndi.org/publications/addressing-online-misogyny-and-gendered-disinformation-how-guide>), as well as a social media analysis tool developed jointly through CEPPS that

Engagement

Facebook has supported the development of Social Science One (<https://socialscience.one/>), a consortium of universities and research groups that have been working to understand various aspects of the online world, including disinformation, hate speech, and computational propaganda. This is also supported by foundations including the John and Laura Arnold Foundation, the Democracy Fund, the William and Flora Hewlett Foundation, the John S. and James L. Knight Foundation, the Charles Koch Foundation, the Omidyar Network, the Sloan Foundation, and Children's Investment Fund Foundation. The project was announced and launched in July 2018. Notably, all but three of the projects are focused on the developed world, and of those three, two projects are in Chile and one in Brazil. Through this consortium, the platform has enabled access to a URLs Data Set of widely shared links (<https://www.poynter.org/fact-checking/2019/these-researchers-are-getting-access-to-facebook-data-to-study-misinformation/>) that is otherwise unavailable to the wider research community.

Facebook received criticism on the program because of the slow speed of implementation, the release and management of research data, and the negotiation of other complicated issues. In all aspects of collaboration with platforms, agreement on data sharing and management are critical components of projects and certainly must be negotiated carefully to avoid sharing private user information. The misuse of such data as

describes

happened during the Cambridge Analytica scandal

(<https://www.fastcompany.com/40550423/how-facebook-blew-it>) should be avoided. The data was later used by private companies to model voter behavior and target advertisements with psychographic and other information from these profiles, creating huge questions about the use of private user data in campaigns and elections. It is important to highlight that the history of this project helped set the terms for research collaboration with Facebook going forward.

([https://www.ifes.org/sites/default/files/violence against women in elections online a social media methodologies for building lexicons in local contexts](https://www.ifes.org/sites/default/files/violence%20against%20women%20in%20elections%20online%20a%20social%20media%20methodologies%20for%20building%20lexicons%20in%20local%20contexts))⁴.

In many cases, there simply are not enough resources to hire even minimal levels of moderators and technologists to deal with what is happening. This creates issues for content moderation, reporting, and algorithmic forms of detection and machine learning to inform these systems. In many cases, moderation efforts are up against information attacks and coordinated inauthentic behavior that go beyond ordinary manipulation and can be sponsored by private or public authorities with deep pockets. In Brazil, WhatsApp's program (<https://www.whatsapp.com/research/awards/announcement/>) supported studies of its election from top researchers in the field. Researchers at the Universities of Syracuse and Minas Gerais studied user information sharing and compared it to voter behavior, while others from the Institute of Technology and Society in Rio looked at methods for training people in media literacy through the platform.

WhatsApp has supported research on the platform and enabled access to its business API in certain cases, such as the First Draft/Comprova (<https://firstdraftnews.org/tackling/comprova/>) project in Brazil. It has also financially supported groups such as the Center for Democracy and Development and the University of Birmingham to pioneer research on the platform in Nigeria.

FEATURED INTERVENTION
ITS RIO CURRICULA ON
INFORMATION ISSUES
(/INTERVENTIONS/ITS-RIO-

CURRICULA-INFORMATION-ISSUES)

The mission of the Institute of Technology and Society (ITS) is to ensure that Brazil and the Global South respond creatively and appropriately to the opportunities provided by technology in the digital age, and that its potential benefits are widely

3. Twitter

Twitter has taken a more comprehensive approach to releasing data than any other company. Since 2018 (https://blog.twitter.com/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter.html), the company has made available comprehensive datasets of state-linked information operations that it has removed. Rather than providing samples or access to only a small number of researchers, Twitter established a public archive (<https://transparency.twitter.com/en/reports/information-operations.html>) of all Tweets and related content that it has removed. The archive now runs into hundreds of millions of Tweets and several terabytes of media.

This archive has enabled a wide range of independent research, as well as collaboration with expert organizations. In 2020, the company partnered with (<https://carnegieendowment.org/specialprojects/counteringinfluenceoperations/workshops>) the Carnegie Partnership for Countering Influence Operations (PCIO) to co-host a series of virtual workshops to support an open exchange of ideas among the research community regarding how IO can be better understood, analyzed, and mitigated. Twitter's API is a unique source of data for the academic community, and the company launched a dedicated academic API (https://blog.twitter.com/developer/en_us/topics/tools/2021/enabling-the-future-of-academic-research-with-the-twitter-api.html) product in 2021.

More broadly, Twitter collaborates frequently with and has provided grants to support a number of organizations working to promote information integrity. Just like Facebook, the company has worked closely with research partners (https://blog.twitter.com/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter.html) like the Stanford Internet Observatory, Graphika, and the Atlantic Council Digital Forensic Research Lab on datasets related to the networks detected and removed from their platform. The platform has also collaborated with the Oxford Internet Institute's Computational Propaganda Project to analyze information operation activities.

4. Microsoft

Microsoft has initiated the Defending Democracy Program, (<https://blogs.microsoft.com/on-the-issues/2018/04/13/announcing-the-defending-democracy-program/>) partnering with various civil society, private sector, and academic groups working on cybersecurity, disinformation, and civic technology issues. As part of this initiative, starting in 2018, Microsoft partnered with Newsguard

(<https://blogs.microsoft.com/on-the-issues/2018/04/13/announcing-the-defending-democracy-program/>), a plug-in to browsers such as Chrome and Edge that validates news websites for users based on nine [journalistic](https://www.newsguardtech.com/ratings/rating-process-criteria/) (<https://www.newsguardtech.com/ratings/rating-process-criteria/>) integrity criteria. Based on this evaluation, the site is given a positive or negative rating, green or red respectively. The plug-in has been downloaded thousands of times, and this technology powers information literacy programs in partnership with libraries and schools.

It has also engaged in research initiatives and partnerships on disinformation, including support for research on [disinformation](https://venturebeat.com/2020/04/07/microsoft-ai-fake-news-better-than-state-of-the-art-baselines/) (<https://venturebeat.com/2020/04/07/microsoft-ai-fake-news-better-than-state-of-the-art-baselines/>) and social media by Arizona State University, the [Oxford Internet Institute](https://oxtec.oii.ox.ac.uk/wp-content/uploads/sites/115/2019/10/OxTEC-The-Market-of-Disinformation.pdf) (<https://oxtec.oii.ox.ac.uk/wp-content/uploads/sites/115/2019/10/OxTEC-The-Market-of-Disinformation.pdf>), [Princeton University's](https://citp.princeton.edu/event/klein-2/) (<https://citp.princeton.edu/event/klein-2/>) Center for Information Technology Policy, as well as [Microsoft Research](https://www.microsoft.com/en-us/research/publication/the-science-of-fake-news/) (<https://www.microsoft.com/en-us/research/publication/the-science-of-fake-news/>) itself.

In a cross-sectoral collaboration, Microsoft, the Bill & Melinda Gates Foundation, and USAID supported the Technology and Social Change group at the University of Washington's Information School to develop a program for [Mobile Information Literacy](https://tascha.uw.edu/mobile-information-literacy-curriculum/) (<https://tascha.uw.edu/mobile-information-literacy-curriculum/>) that includes content verification, search, and evaluation. This project developed into a Mobile Information Literacy Curriculum which has since been applied in Kenya.

5. LINE

[LINE](https://www.poynter.org/fact-checking/2018/how-misinformation-spreads-on-line-%C2%97-one-of-the-most-popular-messaging-apps-in-southeast-asia/) (<https://www.poynter.org/fact-checking/2018/how-misinformation-spreads-on-line-%C2%97-one-of-the-most-popular-messaging-apps-in-southeast-asia/>), as with many other messaging apps, is sometimes taken advantage of by scammers, hoaxers, and fake news writers. While there have not been major claims of systematic disinformation on the platform, [LINE](https://www.poynter.org/fact-checking/2018/how-misinformation-spreads-on-line-%C2%97-one-of-the-most-popular-messaging-apps-in-southeast-asia/) (<https://www.poynter.org/fact-checking/2018/how-misinformation-spreads-on-line-%C2%97-one-of-the-most-popular-messaging-apps-in-southeast-asia/>) has acknowledged issues of false information circulating on its networks. Fact-checkers have developed partnerships with the platform in order to prevent the spread of disinformation, including the CoFacts automated fact-checking system maintained by g0v (pronounced gov zero), a civic technology community in Taiwan. Users can add the [Fact Line Checker](https://restofworld.org/2021/how-line-is-fighting-disinformation-without-sacrificing-privacy/) (<https://restofworld.org/2021/how-line-is-fighting-disinformation-without-sacrificing-privacy/>) to their contacts and forward messages to the checker and receive an answer in real time about whether the content is true or false. This also serves to automatically report suspicious messages to the platform, which allows Line to track misinformation and disinformation without breaking end-to-end encryption.

FEATURED INTERVENTION

COFACTS

(/INTERVENTIONS/COFACTS)

A project of the g0v civic technology community in Taiwan, CoFacts is a fact checking bot for messaging groups. Messages can be forwarded to the CoFacts bot for fact checking by a team of volunteers; the

In September 2019, LINE launched an anti-hoax campaign in partnership with the Associated Press. This [campaign \(https://www.youtube.com/watch?v=6RfHF_OkR9E&feature=youtu.be\)](https://www.youtube.com/watch?v=6RfHF_OkR9E&feature=youtu.be) includes a series of educational videos focused on identifying credible news sources and fake news. In a press release LINE said, "Taking 'Stop Fake News' as the theme, [the campaign \(https://www.youtube.com/watch?v=6RfHF_OkR9E&feature=youtu.be\)](https://www.youtube.com/watch?v=6RfHF_OkR9E&feature=youtu.be) aims to help users improve their media literacy and create a safe digital environment."



HIGHLIGHT

In 2018, a group of international civil society organizations, including IFES, IRI, NDI, and International IDEA, formed [the Design 4 Democracy Coalition \(https://d4dcoalition.org/\)](https://d4dcoalition.org/) to promote coordination among democracy organizations and provide a space for [constructive engagement between the democracy community and technology companies \(/topics/csos/4-advocacy-toward-platforms\)](/topics/csos/4-advocacy-toward-platforms).

B. CROSS-SECTOR AND MULTISTAKEHOLDER INITIATIVES

Increasingly, the major platforms are looking for broader ways to collaborate with civil society, governments, and others to not only combat disinformation, hate speech, and other harmful forms of content on their networks, but also promote better forms of content. These collaborations come in the form of coalitions with different groups, codes of practice, and other joint initiatives.

Facebook, Twitter and other major platforms have, for example, increasingly engaged with research groups such as the [Atlantic Council's](https://about.fb.com/news/2018/05/announcing-new-election-partnership-with-the-atlantic-council/) (<https://about.fb.com/news/2018/05/announcing-new-election-partnership-with-the-atlantic-council/>), Digital Forensic Research (DFR) Lab, Graphika, and others to identify and take down large networks of false or coordinating accounts that are in violation of community standards. In addition, local groups such as [International Society for Fair Elections and Democracy](https://www.isfed.ge/eng/) (<https://www.isfed.ge/eng/>) (ISFED) have also assisted social media platforms with information to facilitate take downs and other enforcement actions. Local organizations are becoming an increasingly important component of the reporting system for various platforms that do not have the capacity to actively monitor and understand local contexts like Georgia.

Among more formal collaborations, the [Global Network Initiative](https://globalnetworkinitiative.org/) (<https://globalnetworkinitiative.org/>) (GNI) dates back to 2005 and continues to support multi-stakeholder engagement among platforms and civil society, particularly on issues related to disinformation and other harmful forms of content. For more information on the GNI, [see the norms and standards chapter](/topics/norms/0-overview-norms) (</topics/norms/0-overview-norms>).

FEATURED INTERVENTION DESIGN 4 DEMOCRACY (/INTERVENTIONS/DESIGN-4- DEMOCRACY)

Our mission is to ensure that information technology and social media play a proactive role in supporting democracy and human rights globally. As a community we create programs, training, and dialogue that promote the safe and responsible use of

Among the cross-sector initiatives to combat disinformation, one of the most prominent is the **European Union's Code of Practice on Disinformation**. The code was developed by a European Union (EU) working group on disinformation . The code supplies member governments and countries that want to trade and work with the bloc guidelines about how to run their regulatory frameworks in line with GDPR and other online EU regulations, as well as plans for responses to disinformation through digital literacy, fact-checking, media, and support for civil society, among other interventions. Based on this code, the EU has developed a [Democracy Action Plan](https://ec.europa.eu/info/strategy/priorities-2019-2024/new-push-european-democracy/european-democracy-action-plan_en), (https://ec.europa.eu/info/strategy/priorities-2019-2024/new-push-european-democracy/european-democracy-action-plan_en) an initiative that the EU plans to implement in the coming year that focuses on promoting free and fair elections, strengthening media freedom, and countering disinformation. Core to its the disinformation efforts are:

- Improving the EU's existing toolbox for countering foreign interference
- Overhauling the Code of Practice on Disinformation into a co-regulatory framework of obligations and accountability of online platforms
- Setting up a robust framework for Code of Practice implementation.

FEATURED INTERVENTION EU CODE OF PRACTICE ON DISINFORMATION (/INTERVENTIONS/EU-CODE- PRACTICE-DISINFORMATION)

The European Union developed a Code of Practice on Disinformation based on the findings of its High Level Working Group on the issue. This included recommendations for companies operating in the EU

At the Internet Governance Forum held at UNESCO in Paris and the Paris Peace Forum in November 2018, the President of the French Republic, Emmanuel Macron, introduced [The Paris Call for Trust and Security in Cyberspace](https://pariscall.international/en/) (<https://pariscall.international/en/>). Signatories (<https://pariscall.international/en/supporters>) to the Call commit to promoting [nine core principles](https://pariscall.international/en/principles) (<https://pariscall.international/en/principles>) and reaffirm various commitments related to international law, cybersecurity, infrastructure protection, and countering disinformation. So far, 79 countries, 35 public authorities, 391 organizations of civil society, and 705 companies and private sector entities have signed on to a common set of principles on stability and security in the

information space. The United States has yet to formally commit or sign on to the initiative. Nevertheless, the initiative represents one of the most ambitious cross-sector collaborations dedicated to cybersecurity and information integrity to date.

RESEARCH TOOLS FOR UNDERSTANDING DISINFORMATION

0. EXECUTIVE SUMMARY (/TOPICS/SURVEYS/0-EXECUTIVE-SUMMARY)

Written by Bret Barrowman, Senior Specialist for Research and Evaluation, Evidence and Learning Practice at the International Republican Institute

Effective democracy, human rights, and governance programming requires practitioners to accurately assess underlying causes of information disorders (<https://www.coe.int/en/web/freedom-expression/information-disorder>) and to evaluate the effectiveness of interventions to treat them. Research serves these goals at several points in the DRG program cycle: problem and context analysis, targeting, design and content development, monitoring, adaptation, and evaluation.

GOALS OF RESEARCH (/TOPICS/SURVEYS/1-OVERVIEW-RESEARCH-TOOLS)

Applying research in the DRG program cycle supports programs by fulfilling the scientific goals of description, explanation, and prediction. Description identifies characteristics of research subjects and general patterns or relationships. Explanation identifies cause and effect relationships. Prediction forecasts what might happen in the future.

RESEARCH FOR CONTEXT ANALYSIS AND DESIGN (/TOPICS/SURVEYS/2-RESEARCH-COUNTER-DISINFORMATION-PROGRAM-DESIGN)

Effective DRG programs to counter disinformation require the identification of a specific problem or set of problems in the information environment in a particular context. Key methods include landscape analysis, stakeholder analysis, political economy analysis, and the use of surveys or interviews to identify potential beneficiaries or particularly salient themes within a specific context.

Sample general research questions:

- What are the main drivers of disinformation in this context?
- What are the incentives for key actors to perpetuate or mitigate disinformation in this context?
- Through which medium is disinformation likely to have the greatest impact in this context?
- What evidence suggests our proposed activity(ies) will mitigate the problem?
- Which groups are the primary targets or consumers of disinformation in this context?
- Which key issues or social cleavages are most likely to be subjects of disinformation in this context?

IMPLEMENTATION RESEARCH (/TOPICS/SURVEYS/3-RESEARCH-COUNTER-DISINFORMATION-PROGRAM-IMPLEMENTATION)

There are several research and measurement approaches available for practitioners to monitor activities related to information and disinformation, both for program accountability functions and for adaptation to changing conditions. Key methods include digital and analog media audience metrics, measurement of knowledge, attitudes, or beliefs with surveys or focus groups, media engagement metrics, network analysis, and A/B tests. Key research questions include:

- How many people are engaging in program activities or interventions?
- What demographic, behavioral, or geographic groups are engaging in program activities? Is the intervention reaching its intended beneficiaries?
- How are participants, beneficiaries, or audiences reacting to program activities or materials? How do these reactions differ across subgroups, and specifically marginalized groups?
- Is one mode or message more effective than another in causing audience to engage information and/or share it with others? How does information uptake and sharing differ across subgroups? What are barriers to information or program uptake among marginalized groups?
- What framing of content is most likely to reduce consumption of disinformation, or increase consumption of reliable information? For example, is a fact-checking message more likely to cause consumers to update their beliefs in the direction of truth, or does it cause retrenchment in belief in the original disinformation? Does this effect vary across subgroups?

EVALUATION RESEARCH (/TOPICS/SURVEYS/4-EVALUATIVE-RESEARCH-COUNTER-DISINFORMATION-PROGRAMS)

DRG program and impact evaluation can identify and describe key results, assess or improve the quality of program implementation, identify lessons that might improve the implementation of similar programs, or attribute changes in key outcome to a program intervention. Key methods include randomized evaluations and quasi- or non-experimental evaluations, including pre/post designs, difference-in-differences, statistical matching, comparative case studies, process tracing, and regression analysis. Key research questions include:

- Are there observable outcomes associated with the program?
- Does a program or activity cause a result of interest? For example, did a media literacy program increase the capacity of participants to distinguish between true news and false news? Does a program cause unintended outcomes?
- What is the size of the effect (i.e., impact) of an activity on an outcome of interest?
- What is the direction of the effect of an activity on an outcome of interest? For example, did a fact checking program decrease confidence in false news reports, or did it cause increased acceptance of those reports through backlash?

RECOMMENDATIONS (/TOPICS/SURVEYS/5-RECOMMENDATIONS)

- Specific research questions should drive the selection of research designs and data collection methods. Committing to a specific design or data collection method will limit the questions the researcher is able to answer.
- Use a pilot-test-scale model for program activities or content. Using one or more of these research approaches, workshop interventions on small groups of respondents, and use pilot data to refine promising approaches before deploying to a larger set of beneficiaries.
- Protect personally identifiable information (PII). All the data collection methods described in this section can collect information characteristics, attitudes, beliefs, and willingness to engage in political action. Regardless of the method, researchers should make every attempt to secure informed consent to participate in research and should take care to secure and de-identify personal data.
- Consider partnerships with research organizations, university labs, or individual academic researchers, who may have a comparative advantage in designing and implementing complex research designs, and who may have an interest in studying the effects of counter-disinformation programs.

RESEARCH TOOLS FOR UNDERSTANDING DISINFORMATION

1. OVERVIEW - RESEARCH TOOLS (/TOPICS/SURVEYS/1-OVERVIEW-RESEARCH-TOOLS)

Effective democracy, human rights, and governance (DRG) programming to respond to disinformation requires practitioners to make accurate inferences about the underlying causes of information disorders and about the effects of their interventions. Programs to counter disinformation often rely on a research component to identify problems, to identify potential targets or beneficiaries of an intervention, to develop and adapt program content, to monitor implementation, and to evaluate results. This chapter will survey a broad menu of research tools and approaches for understanding disinformation and potential responses, with the goal of supporting DRG practitioners in designing, implementing, and evaluating programs based on the best available data and evidence.

The sections that follow distinguish broadly between research approaches or designs and data collection methods.

To support DRG practitioners in developing evidence-based programs to counter disinformation, this chapter is structured according to stages in the program cycle – design, implementation, and evaluation. It provides examples of research approaches that can help answer questions for specific decisions at each stage. As a final note, the examples provided are suggestive, not exhaustive. Useful and interesting research and data collection methods, especially on information and disinformation, require thought, planning, and creativity. To develop a research approach that is most useful for a program, consider consulting or partnering early with internal experts including applied researchers and evaluators or external experts



HIGHLIGHT

For the purposes of this guide, a research approach or research design refers to a method or set of methods that allow researchers or practitioners to make valid inferences about disinformation or programmatic responses. In other words, a research design is a method through which one can confidently and accurately answer specific research questions. On the other hand, data collection describes the ways in which researchers and practitioners collect the information needed to answer those

through one of many academic institutions that specialize in research on democracy and governance interventions.

research questions. For example, key informant interviews (KIIs) or in-depth interviews (IDIs) are data collection methods that may be used within several research designs.

RESEARCH NETWORKS

EGAP (<https://egap.org/>): Evidence in Governance and Politics (EGAP) is a research, evaluation, and learning network with worldwide reach that promotes rigorous knowledge accumulation, innovation, and evidence-based policy in various governance domains, including accountability, political participation, mitigation of societal conflict, and reducing inequality. It does so by fostering academic-practitioner collaborations, developing tools and methods for analytical rigor, and training academics and practitioners alike, with an intensive focus in the Global South. Results from research are shared with policy makers and development agencies through regular policy fora, thematic and plenary meetings, academic practitioner events, and policy briefs.

J-PAL (<https://www.povertyactionlab.org/>): The Abdul Latif Jameel Poverty Action Lab (J-PAL) is a global research center working to reduce poverty by ensuring that policy is informed by scientific evidence. Anchored by a network of 227 affiliated professors at universities around the world, J-PAL conducts randomized impact evaluations to answer critical questions in the fight against poverty. J-PAL translates research into action, promoting a culture of evidence-informed policymaking around the world. Their policy analysis and outreach help governments, NGOs, donors, and the private sector apply evidence from randomized evaluations to their work and contributes to public discourse around some of the most pressing questions in social policy and international development.

IPA (<http://www.poverty-action.org/>): Innovations for Poverty Action (IPA) is a research and policy nonprofit that discovers and promotes effective solutions to global poverty problems. IPA brings together researchers and decision-makers to design, rigorously evaluate, and refine these solutions and their applications, ensuring that the evidence created is used to improve the lives of the world's poor.

Political Violence FieldLab (<https://dickey.dartmouth.edu/programs/security/political-violence-fieldlab>): The Political Violence FieldLab provides a home for basic and applied research on the causes and effects of political violence. The FieldLab provides students the opportunity to work on cutting-edge and policy-relevant questions in the study of political violence. Their projects involve close collaboration with government agencies and non-government organizations to evaluate the effects and effectiveness of interventions in contemporary conflict settings.

MIT GovLab (<https://mitgovlab.org/>): GovLab collaborates with civil society, funders, and governments on research that builds and tests theories about how innovative programs and interventions affect political behavior and make governments more accountable to citizens. They develop and test hypotheses about accountability and citizen engagement that contribute to theoretical knowledge and help practitioners learn in real time. Through integrated and sustained collaborations, GovLab works together with practitioners at every stage of the research, from theory building to theory testing.

DevLab@Duke (<https://www.devlabduke.com/>): The DevLab@Duke is an applied learning environment that focuses on connecting social scientists at Duke who work in international development with the community of development practitioners to create rigorous programming, collect monitoring and evaluation data, and conduct impact evaluations of development projects. In addressing these goals, they bring together scholars and students attuned to the research frontier and with advanced capabilities in experimental and quasi-experimental impact evaluation designs, survey design and other data collection tools, and data analytics, including impact evaluation econometrics, web scraping and geospatial analysis.

Center for Effective Global Action (CEGA) (<https://cega.berkeley.edu/>): CEGA is a hub for research on global development. Headquartered at the University of California, Berkeley, their large, interdisciplinary network—including a growing number of scholars from low and middle-income countries—identifies and tests innovations designed to reduce poverty and promote development. CEGA researchers use rigorous evaluations, tools from data science, and new measurement technologies to assess the impacts of large-scale social and economic development programs.

Citizens and Technology Lab (<https://citizensandtech.org/>): Citizens and Technology Lab does citizen science for the internet. They seek to enable anyone to engage critically with the tech tools and platforms they use, ask questions, and get answers. Working hand-in-hand with diverse communities and organizations around the world, they identify issues of shared concern (“effects”) related to digital discourse, digital rights and consumer protection. Their research methods can discover if a proposed effect is really happening, uncover the causes behind a systemic issue, and test ideas for creating change.

Stanford Internet Observatory (<https://cyber.fsi.stanford.edu/io>): The Stanford Internet Observatory is a cross-disciplinary program of research, teaching, and policy engagement for the study of abuse in current information technologies, with a focus on social media. The Observatory was created to learn about the abuse of the internet in real time, to develop a novel curriculum on trust and safety that is a first in computer science, and to translate research discoveries into training and policy innovations for the public good.

GOALS OF RESEARCH

Description, Explanation, or Prediction? (<https://psychcentral.com/blog/understanding-research-methodology-3-goals-of-scientific-research#1>) Applied research in the DRG program cycle can

support programs by fulfilling one or more of the following scientific goals.

Description: Descriptive research aims to identify characteristics of research subjects at different levels of analysis (e.g., individual, group, organization, country, etc.). Descriptive research classifies or categorizes subjects or identifies general patterns or relationships. Examples of descriptive research in countering disinformation programs might include developing descriptive statistics in polling or survey data to identify key target groups, or analysis to identify key themes in media content.

Explanation: Explanatory research aims to identify cause and effect relationships; it helps answer “why?” questions. It establishes causation through sequencing (as causes must precede their effects) and/or eliminating competing explanations through comparisons. This category may also include evaluation research in the program cycle, to the extent an evaluation attempts to determine the “impact” of a program on an outcome of interest (i.e., whether a program causes a result), or to determine which of several potential program approaches is most effective.

Prediction: Predictive research uses descriptive or explanatory methods to forecast what might happen in the future. At a basic level, predictive research in the DRG program cycle might involve using findings from a program evaluation to adapt approaches to the next cycle or to another context. More systematic predictive research uses qualitative or quantitative methods to assign specific probabilities to events over a designated time, as in a weather forecast.

Data sources and collection methods for Disinformation Research include Key Informant Interviews (KII), Focus Groups, Public Opinion Polls, Surveys, Audience Metrics (analog and digital), Web and Social Media Scraping, Administrative Data analysis ([data collected and stored as part of the operations of organizations like governments, nonprofits, or firms](https://www.povertyactionlab.org/blog/9-25-20/announcing-handbook-using-administrative-data-research-and-evidence-based-policy) (<https://www.povertyactionlab.org/blog/9-25-20/announcing-handbook-using-administrative-data-research-and-evidence-based-policy>)). There are other methods but these are some key ones that will be explored further in this text.



HIGHLIGHT

Predictive Research in DRG Programming

Several USAID-funded initiatives use predictive research to help DRG practitioners better anticipate and respond to changes in political context. For example, the CEPPS Democratic Space Barometer forecasts democratic opening and closing over a two-year window. The Internews-led INSPIRES Consortium uses media scraping and machine learning to forecast closing civic space on a monthly basis.

RESEARCH TOOLS FOR UNDERSTANDING DISINFORMATION

2. RESEARCH FOR COUNTER-DISINFORMATION PROGRAM DESIGN (/TOPICS/SURVEYS/2-RESEARCH-COUNTER-DISINFORMATION-PROGRAM-DESIGN)

Practitioners must make several key decisions in the counter-disinformation program design phase. Those decisions include identifying a specific set of problems the program will address, developing a logic through which the program will address that problem, selecting between alternative activities, and deciding who will be the primary targets or beneficiaries of those activities.

CONTEXT ANALYSIS AND PROBLEM STATEMENTS

Effective DRG programs to counter disinformation require the identification of a specific problem or set of problems in the information environment in a particular context.

DRG practitioners rely on several research methods to identify priority issues, context-specific drivers of information disorders, perpetrators and targets of disinformation, and incentives to perpetuate or mitigate disinformation. [Landscape and stakeholder analyses](#)



HIGHLIGHT

Tool spotlight: [Hewlett Foundation Literature Review](https://hewlett.org/library/social-media-political-polarization-political-disinformation-review-scientific-literature/) (https://hewlett.org/library/social-media-political-polarization-political-disinformation-review-scientific-literature/)

“The Hewlett Foundation commissioned this report to provide an overview of the current state of the literature on the relationship between social media; political polarization; and political “disinformation,” a term used to encompass a wide range of types of information about politics found online, including “fake news,” rumors, deliberately factually incorrect information,

inadvertently factually incorrect information, politically slanted information, and “hyperpartisan” news.

The review of the literature is provided in six separate sections, each of which can be read individually but that cumulatively are intended to provide an overview of what is known—and unknown—about the relationship between social media, political polarization, and disinformation. The report concludes by identifying key research gaps in our understanding of these phenomena and the data that are needed to address them.”

(<http://www1.worldbank.org/publicsector/anticorrupt/PoliticalEconomy/stakeholderreading.htm>) are approaches to answer key **descriptive** research questions about the information environment, including identifying important modes of communication, key media outlets, perpetrators and target audiences for disinformation, and key political issues or personalities that might be the subjects of disinformation. Of note, women and members of other marginalized groups have been victims of political and sexualized disinformation, online hate, and harassment. As such, DRG practitioners should also account for uniquely targeted disinformation aimed at marginalized populations (<https://staging.counterdisinformation.org/topics/gender/1-gender-considerations-counter-disinformation-programming>) globally by conducting qualitative, quantitative, and gender sensitive, inclusive research in order to understand these important dynamics.

These methods may also be **explanatory**, inasmuch as they identify key causes or drivers of specific information disorders.

As an exploratory option, key data collection methods often include key informant interviews (KII) with respondents identified through convenience or snowball sampling



HIGHLIGHT

Sample general research questions:

- What are the main drivers of disinformation in this context?
- What are the incentives for key actors to perpetuate or mitigate

- disinformation in this context?
- Through which medium is disinformation likely to have the greatest impact in this context?
- What evidence suggests our proposed activity(ies) will mitigate the problem?
- What groups are the primary targets or consumers of disinformation in this context?
- What key issues or social cleavages are most likely to be subjects of disinformation in this context?

https://usaidlearninglab.org/sites/default/files/resource/files/sampling_design_review_final_kuzara
Surveys and public opinion polls can also be valuable tools for understanding the media and information landscape. Survey questionnaire items on the media landscape can inform programming by identifying how most people get news on social or political events, what outlets are most popular among specific demographic or geographic groups, or which social or political issues are particularly polarizing. Respondents for surveys or polls, if possible, should be selected via a method of sampling that eliminates potential selection biases

https://usaidlearninglab.org/sites/default/files/resource/files/sampling_design_review_final_kuzara
to ensure that responses are representative of a larger population of interest. Landscape and stakeholder analyses may also rely on desk research on primary and secondary sources, such as state administrative data (e.g. census data, media ownership records, etc.), journalistic sources like news or investigative reports, academic research, or program documents from previous or ongoing programs.

Applied Political Economy Analysis (PEA) (<https://www.usaid.gov/documents/1866/thinking-and-working-politically-through-applied-political-economy-analysis>) is a contextual research approach that focuses on identifying the incentives and constraints that shape the decisions of key actors in an information environment. This approach goes beyond technical solutions to information disorders to analyze why and how key actors might perpetuate or mitigate disinformation, and subsequently, how these social, political, or cultural factors may affect the implementation, uptake, or impact of programmatic responses. Like other context analysis approaches, PEA relies on both existing research gathered and analyzed through desk review and data collection of experiences, beliefs, and perceptions of key actors.

RESEARCH TOOLS FOR UNDERSTANDING

DISINFORMATION

3. RESEARCH FOR COUNTER-DISINFORMATION PROGRAM IMPLEMENTATION

(/TOPICS/SURVEYS/3-RESEARCH-COUNTER-DISINFORMATION-PROGRAM-IMPLEMENTATION)

There are several research and measurement tools available to assist practitioners in monitoring of activities related to information and disinformation. At a basic level, these tools support program and monitoring, evaluation, and learning (MEL) staff in performing an accountability function. However, these research tools also play an important role in adapting programming to changing conditions. Beyond answering questions about whether and to what extent program activities are engaging their intended beneficiaries, these research tools can help practitioners identify how well activities or interventions are performing so that implementers can iterate, as in an [adaptive management \(https://usaidlearninglab.org/lab-notes/what-adaptive-management-0\)](https://usaidlearninglab.org/lab-notes/what-adaptive-management-0) or [Collaborating, Learning, and Adapting \(CLA\) \(https://usaidlearninglab.org/qrg/understanding-cla-0\)](https://usaidlearninglab.org/qrg/understanding-cla-0) framework.

Program Monitoring

(assess implementation, if content is reaching desired targets, if targets are engaging content)

Key Research Questions:

- How many people are engaging in program activities or interventions?
- What demographic, behavioral, or geographic groups are engaging in program activities? Is the intervention reaching its intended beneficiaries?
- How are participants, beneficiaries, or audiences reacting to program activities or materials?
- How does engagement or reaction vary across activity types?

Several tools are available to assist DRG practitioners in monitoring the reach of program activities and the degree to which audiences and intended beneficiaries are engaging program content. These tools differ according to the media through which information and disinformation, as well as counter-programming, are distributed. For analog media outlets like television and radio, audience metrics, including size, demographic composition, and geographic reach may be available through the outlets themselves or through state administrative records. The usefulness and detail of this information depends on the capacity of the outlets to collect this information and their willingness to share it publicly. Local marketing or advertising firms may also be good

sources of audience information. In some cases, the reach of television and/or radio may be modeled (<https://onlinelibrary.wiley.com/doi/full/10.1111/ajps.12355>) using information on the broadcast infrastructure.

Digital platforms provide a more accessible suite of metrics. Social media platforms like Twitter, Facebook, and YouTube have built in analytical tools that allow even casual users to monitor post views engagements (including “likes,” shares, and comments). Depending on the platform Application Programming Interface (API) and terms of service, more sophisticated analytical tools may be available. For example, Twitter’s API allows users to import large volumes of both metadata and tweet content, enabling users to monitor relationships between accounts and conduct content or sentiment analysis around specific topics. Google Analytics (<https://support.google.com/analytics/#topic=9143232>) provides a suite of tools for measuring consumer engagement with advertising material, including behavior on destination websites. For example, these tools can help practitioners understand how audiences, having reached a resource or website by clicking on digital content (e.g. links embedded in tweets, Facebook posts, or YouTube video) are spending time on the destination resources and what resources they are viewing, downloading, or otherwise engaging. Tracking click-throughs provides potential measures of destination behavior, not just beliefs or attitudes.

Workshopping Content: Pilot-Test-Scale

Determining the content of programmatic activities is a key decision point in any program cycle. With respect to counter-disinformation programs, implementers should consider how the messenger, mode, and content of an intervention is likely to influence uptake and engagement by target groups with that content, and whether the material is likely to change beliefs or behavior. With this in mind, workshopping and testing counter-disinformation content throughout the implementation program phase can help implementers identify which programmatic approaches are working, as well as how and whether to adapt content in response to changing conditions.

Key Research Questions:

- What modes or messengers are most likely to increase content uptake in this context? For example, is one approach more effective than another in causing the interpreters to engage information and/or share it with others?
- What framing of content is most likely to reduce consumption of disinformation, or increase consumption of true information in this context? For example, is a fact-checking message more likely to cause consumers to update their beliefs in the direction of truth, or does it cause retrenchment in belief in the original disinformation?

Several data collection methods allow DRG practitioners to workshop the content of interventions with small numbers of potential beneficiaries before scaling activities to larger audiences. Focus groups (scientifically sampled, structured, small group discussions) are used regularly both in market research and DRG programs to elicit in-depth reactions to test products. This format allows researchers to observe spontaneous reactions to prompts and probe respondents for

more information, as opposed to surveys, which may be more broadly representative, but rely on respondents selecting uniform and predetermined response items that do not capture as much nuance. Focus groups are useful for collecting initial impressions about a range of alternatives for potential program content before scaling activities to a broader audience.

A/B tests are a more rigorous method for determining what variations in content or activities are most likely to achieve desired results, especially when alternatives are similar and differences between them are likely to be small. A/B tests are a form of randomized evaluation in which a researcher randomly assigns members of a pool of research participants to receive different versions of content. For example, product marketing emails or campaign fundraising solicitations might randomly assign a pool of email addresses to receive the same content under one of several varying email subjects. Researchers then measure differences between each of these experimental groups on the same outcomes, which for digital content often includes engagement rates, click-throughs, likes, shares, and/or comments.

Social media platforms have used A/B testing to optimize platform responses to misinformation. In other cases, researchers or technology companies themselves have experimented with variations of political content labels



HIGHLIGHT

Mode: The mechanisms through which programmatic content is delivered (e.g. in person, written materials, television, radio, social media, email, SMS, etc.)



HIGHLIGHT

Because participants are randomly assigned to receive different variations, the researcher can confidently conclude any differences over these outcomes can be attributed to the content variation.

(<https://misinforeview.hks.harvard.edu/article/state-media-warning-labels-can-counteract-the-effects-of-foreign-misinformation/>) to determine whether these tags affect audience engagement. Similarly, DRG programs might use A/B testing to optimize digital content on disinformation programs to explore, for instance, how different framings or endorsers of fact-checking messages affect audience beliefs.

Dummy text



HIGHLIGHT

Tools Spotlight: Content and Message Testing Tools

Facebook

(<https://www.facebook.com/business/help/1738?id=445653312788501>): "A/B testing lets you change variables

(<https://www.facebook.com/business/help/1962>

such as your ad creative, audience, or placement to determine which strategy performs best and improve future campaigns. For example, you might hypothesize

(<https://www.facebook.com/business/help/2358>

a custom audience strategy will outperform an interest-based audience strategy for your business. An A/B test lets you quickly compare both strategies to see which one performs best."

RIWI (<https://riwi.com/market/private-enterprise/>): "Respondents are randomly

assigned to a treatment or control group to determine the impact of different concepts, videos, ads or phrases. All groups will see identical initial questions, followed by treatment group(s) receiving a developed message. After the treatment, all respondents will be asked questions to

determine the resonance and engagement of the message or to measure behavioral changes (assessed post-treatment) between groups.”

GeoPoll: (<https://www.geopoll.com/concept-testing/>). “GeoPoll works with leading global brands to test new concepts through video and picture surveys and mobile-based focus groups. Using GeoPoll’s research capabilities and large panel of respondents, brands can reach their target audience and gather much-needed data on what messaging is most effective, how new products should be marketed, how consumers will react to new products, and more.”

Mailchimp
(<https://mailchimp.com/help/about-ab-testing-campaigns/>): “A/B testing campaigns test different versions of a single email to see how small changes can have an impact on your results. Choose what you want to test, like the subject line or content, and compare results to find out what works and what doesn't work for your audience.”

RESEARCH TOOLS FOR UNDERSTANDING DISINFORMATION

4. EVALUATIVE RESEARCH FOR COUNTER-DISINFORMATION PROGRAMS (/TOPICS/SURVEYS/4-EVALUATIVE-

RESEARCH-COUNTER-DISINFORMATION-PROGRAMS)

Evaluation of DRG programs can identify and describe key results, assess or improve the quality of program implementation, identify lessons that might improve the implementation of similar programs, or attribute changes in key outcomes to a program intervention. This section generally focuses on the last type of evaluation– impact evaluation (https://www.betterevaluation.org/en/themes/impact_evaluation), or determining the extent to which a program contributed to changes in outcomes of interest.

Attributing observed results to programs is perhaps the most difficult research challenge in the DRG program cycle. However, there are several evaluation research designs that can help DRG practitioners determine whether programs have an effect on an outcome of interest, whether programs cause unintended outcomes, which of several alternatives is more likely to have had an effect, whether that effect is positive or negative, and how large that effect might be. Often, these methods can be used within the program cycle to optimize activities, especially within a CLA (<https://usaidlearninglab.org/faq/collaborating%2C-learning%2C-and-adapting-cla>), adaptive management (<https://usaidlearninglab.org/lab-notes/what-adaptive-management-0>), or pilot-test-scale (<https://www.usaid.gov/div>) framework.

Programs to counter disinformation can take many forms with many possible intended results, ranging from small-scale trainings of journalists or public officials, to broader media literacy campaigns, to mass communications such as fact-checking or rating media outlets. There is no one-size-fits-all evaluation research approach that will work for every disinformation intervention. DRG program designers and implementers should consider consulting with internal staff and applied researchers, external evaluators, or academic researchers to develop an evaluation approach that answers research questions of interest to the program, accounting for practical constraints in time, labor, budget, scale, and M&E capacity.

Key Research Questions:

- Does a program or activity cause a measurable change in an outcome of interest? For example, did a media literacy program increase the capacity of participants to distinguish between true news and false news? Does a program cause unintended outcomes?
- What is the size of the effect or impact of an activity on an outcome of interest?
- What is the direction of the effect of an activity on an outcome of interest? For example, did a fact checking program decrease confidence in false news reports, or did it cause increased acceptance of those reports through backlash?

Randomized or Experimental Approaches

Randomized evaluations (also commonly called randomized controlled trials (RCTs) or field experiments) are often referenced as the gold standard for **causal inference** – determining whether and how an intervention caused an outcome of interest. Where they are feasible logistically, financially, and ethically, RCTs are the best available method for causal inference because they control for **confounding variables** – factors other than the intervention that might

have caused the observed outcome. RCTs control for these alternative explanations by randomly assigning participants to one or more “treatment” groups (in which they receive a version of the intervention in question) or a “comparison” or “control” group (in which participants receive no intervention or placebo content.) Since participants are assigned randomly to treatment or control, any observed differences in outcomes between those groups can be attributed to the intervention itself. In this way, RCTs can help practitioners and researchers estimate the effectiveness of an intervention.

The costs and logistical commitments for a randomized impact evaluation can be highly variable, depending in large part on the costs of outcome data collection. However, informational interventions, including those intended to counter disinformation, may be particularly amenable to randomized evaluations, as digital tools can support less expensive data collection than face to face methods like interviews or in-person surveys. Regardless of data collection methods, however, randomized evaluations require significant technical expertise and logistical planning, and will not be appropriate for every program, especially those that operate at relatively small scale, since randomized evaluations require large numbers of units of observation in order to identify statistically significant differences. . These evaluation approaches should not be used to evaluate every program. Other impact evaluation methods differ in how they approximate randomization to measure the effect of interventions on observed outcomes, and may be more appropriate for certain program designs.

In 2020, RAND Corporation researchers, in partnership with IREX's [Learn2Discern](https://www.irex.org/news/randomized-control-trial-finds-irexs-media-literacy-messages-be-effective-reducing-engagement) (<https://www.irex.org/news/randomized-control-trial-finds-irexs-media-literacy-messages-be-effective-reducing-engagement>) program in Ukraine, conducted a randomized control trial to estimate both the impact of a Russian disinformation campaign and of a programmatic response that included content labeling and media literacy interventions. The experiment found that Russian propaganda produced emotional reactions and social media engagement among strong partisans, but that those effects were mitigated by labeling the source of the content, and by showing recipients a short video on media literacy.

Quasi-Experimental and Non-Experimental Approaches



HIGHLIGHT

For a comprehensive guide on using randomized evaluations for **causal inference**

(<https://www.povertyactionlab.org/research-resources?view=toc>) in development programming, see J-PAL's Research Resources.



HIGHLIGHT

Researchers and evaluators may employ quasi-experimental or non-experimental approaches when random assignment to treatment and control is impractical or unethical. As the name suggests, these research designs attempt to attribute changes in outcomes to interventions by approximating random assignment to treatment and control conditions through comparisons. In most cases, this approximation involves collecting data on a population that did not participate in a program, but which is plausibly similar to program participants in other respects. Perhaps the most familiar of these methods for DRG practitioners is a pre-/post-test design, in which program participants are surveyed or tested on the same set of questions both prior to and following their participation in the program. For example, participants in a media literacy program might take a quiz that asks them to distinguish between true and false news, both before and after their participation in the program. In this case, the pre-test measures the capacity of an approximation of a “control” or “comparison” group, and the post-test measures that capacity in a “treatment” group of participants who have received the program. Any increase in the capacity to distinguish true and false news is attributed to the program. Structured comparative case studies and process-tracing are examples of non-experimental designs that control for confounding factors through across-case comparisons or through comparison within the same case over time.

Research Spotlight: [Russian Propaganda Hits Its Mark: Experimentally Testing the Impact of Russian Propaganda and Counter-Interventions](https://www.rand.org/pubs/research_reports/R3.html) (https://www.rand.org/pubs/research_reports/R3.html).

There are a variety of quasi-experimental and observational research methods available for program impact evaluation. The choice of these tools to evaluate the impact of a program depends on available data (or capacity to collect necessary data) and the assumptions that are required to identify reliable estimates of program impact. This table, reproduced in its entirety with the written consent of the Abdul Latif Jameel Poverty Action Lab, provides a menu of these options with their respective data collection requirements and assumptions.

METHOD	DESCRIPTION	WHAT ASSUMPTIONS ARE REQUIRED, AND HOW DEMANDING ARE THE ASSUMPTIONS?	REQUIRED DATA
--------	-------------	---	---------------

Randomization	Randomized Evaluation/ Randomized Control Trial	Measure the differences in outcomes between randomly assigned program participants and non-participants after the program took effect.	The outcome variable is only affected by program participation itself, not by assignment to participate in the program or by participation in the randomized evaluation itself. Examples for such confounding effects could be information effects, spillovers, or experimenter effects. As with other methods, the sample size needs to be large enough so that the two groups are statistically comparable; the difference being that the sample size is chosen as part of the research design.	Outcome data for randomly assigned participants and non-participants (the treatment and control groups).
---------------	--	--	---	--

Basic non-experimental comparison methods	Pre-Post	Measure the differences in outcomes for program participants before the program and after the program took effect.	There are no other factors (including outside events, a drive to change by the participants themselves, altered economic conditions, etc.) that changed the measured outcome for participants over time besides the program. In stable, static environments and over short time horizons, the assumption might hold, but it is not possible to verify that. Generally, a diff-in-diff or RDD design is preferred (see below).	Data on outcomes of interest for program participants before program start and after the program took effect.
---	----------	--	---	---

	Simple Difference	Measure the differences in outcomes between program participants after the program took effect and another group who did not participate in the program.	There are no differences in the outcomes of participants and non-participants except for program participation, and both groups were equally likely to enter the program before it started. This is a demanding assumption. Nonparticipants may not fulfill the eligibility criteria, live in a different location, or simply see less value in the program (self-selection). Any such factors may be associated with differences in outcomes independent of program participation. Generally, a diff-in-diff or RDD design is preferred (see below).	Outcome data for program participants as well as another group of nonparticipants after the program took effect.
--	-------------------	--	---	--

Differences in Differences

Measure the differences in outcomes for program participants before and after the program relative to nonparticipants.

Any other factors that may have affected the measured outcome over time are the same for participants and non-participants, so they would have had the same time trajectory absent the program. Over short time horizons and with reasonably similar groups, this assumption may be plausible. A "placebo test" can also compare the time trends in the two groups before the program took place. However, as with "simple difference," many factors that are associated with program participation may also be associated with outcome changes over time. For example, a person who expects a large improvement in the near future may not join the

Data on outcomes of interest for program participants as well as another group of nonparticipants before program start and after the program took effect.

		program (self-selection).	
--	--	---------------------------	--

More nonexperimental methods

Multivariate Regression/OLS

The “simple difference” approach can be— and in practice almost always is— carried out using multivariate regression. Doing so allows accounting for other observable factors that might also affect the outcome, often called “control variables” or “covariates.” The regression filters out the effects of these covariates and measures differences in outcomes between participants and nonparticipants while holding the effect of the covariates constant.

Besides the effects of the control variables, there are no other differences between participants and non-participants that affect the measured outcome. This means that any unobservable or unmeasured factors that do affect the outcome must be the same for participants and nonparticipants. In addition, the control variables cannot in any way themselves be affected by the program. While the addition of covariates can alleviate some concerns with taking simple differences, limited available data in practice and unobservable factors mean that the method has similar issues as simple difference (e.g., self-selection).

Outcome data for program participants as well as another group of non-participants, as well as “control variables” for both groups.

	Statistical Matching	<p>Exact matching: participants are matched to non-participants who are identical based on “matching variables” to measure differences in outcomes. Propensity score matching uses the control variables to predict a person’s likelihood to participate and uses this predicted likelihood as the matching variable.</p>	<p>Similar to multivariable regression: there are no differences between participants and non-participants with the same matching variables that affect the measured outcome. Unobservable differences are the main concern in exact matching. In propensity score matching, two individuals with the same score may be very different even along observable dimensions. Thus, the assumptions that need to hold in order to draw valid conclusions are quite demanding.</p>	<p>Outcome data for program participants as well as another group of non-participants, as well as “matching variables” for both groups.</p>
--	----------------------	---	--	---

Regression
Discontinuity
Design (RDD)

In an RDD design, eligibility to participate is determined by a cutoff value in some order or ranking, such as income level. Participants on one side of the cutoff are compared to non-participants on the other side, and the eligibility criterion is included as a control variable (see above).

Any difference between individuals below and above the cutoff (participants and non-participants) vanishes closer and closer to the cutoff point. A carefully considered regression discontinuity design can be effective. The design uses the “random” element that is introduced when two individuals who are similar to each other according to their ordering end up on different sides of the cutoff point. The design accounts for the continual differences between them using control variables. The assumption that these individuals are similar to each other can be tested with observables in the data. However, the design limits the

Outcome data for program participants and non-participants, as well as the “ordering variable” (also called “forcing variable”).

			comparability of participants further away from the cutoff.	
	Instrumental Variables	The design uses an “instrumental variable” that is a predictor for program participation. The method then compares individuals according to their predicted participation, rather than actual participation.	The instrumental variable has no direct effect on the outcome variable. Its only effect is through an individual’s participation in the program. A valid instrumental variable design requires an instrument that has no relationship with the outcome variable. The challenge is that most factors that affect participation in a program for otherwise similar individuals are also in some way directly related to the outcome variable. With more than one instrument, the assumption can be tested.	Outcome data for program participants and non-participants, as well as an “instrumental variable.

Note. From Sautmann, Anja, and Abdul Latif Jameel Poverty Action Lab (J-PAL). 2019. "Impact evaluation methods" J-PAL Publication. Last Modified 2020

(<https://www.povertyactionlab.org/resource/introduction-randomized-evaluations>)

Media Monitoring and Content Analysis

Media monitoring and content analysis approaches generally aim to answer research questions about whether, how, or why interventions change audience engagement with information or the nature or quality of the information itself. For example, a fact-checking program might hypothesize that correcting disinformation should result in less audience engagement with outlets for disinformation on social media, as measured by views, likes, shares, or comments.

Several tools are available to help DRG practitioners and researchers identify changes in media content. Content analysis is a qualitative research approach through which researchers can identify key themes in written, audio, or video material, and whether those themes change over time. Similarly, sentiment analysis can help identify the nature of attitudes or beliefs around a theme.

Both content and sentiment analysis can be conducted using human or machine-assisted coding and should be conducted at multiple points in the program cycle in conjunction with other evaluation research designs for project impact evaluation.

Network Analysis

Network analysis is a method for understanding how and why the structure of relationships between actors affects an outcome of interest. Network analysis is a particularly useful research method for countering disinformation programs because it allows analysts to visualize and understand how information is disseminated through online networks, including social media platforms, discussion boards, and other digital communities. By synthesizing information on the number of actors, the frequency of interactions between actors, the quality or intensity of interactions, and the structure of relationships, network analysis can help researchers and practitioners identify key channels for the propagation of disinformation, the direction of transmission of information or disinformation, clusters denoting distinct informational ecosystems, and whether engagement or amplification is genuine or artificial. In turn, network metrics can help



HIGHLIGHT

Research Spotlight: IREX Learn2Discern Quasi-Experimental Impact Evaluation

From October 2015 to March 2016, IREX Implemented Learn2Discern – a large-scale media literacy program in Ukraine in collaboration with The Academy of Ukrainian Press and StopFake. As part of the program, IREX conducted a quasi-experimental impact evaluation using statistical matching to compare program participants to non-participants. The study found that program participants were:

- 28% more likely to demonstrate sophisticated knowledge of the news media industry

inform the design, content, and targeting of program activities

- 25% more likely to self-report checking multiple news sources
- 13% more likely to correctly identify and critically analyze a fake news story
- 4% more likely to express a sense of agency over what news sources they can access.

Donors and partners implementing countering disinformation programs should consider these quasi-experimental methods to evaluate the direction and magnitude of program impacts on outcomes of interest, particularly where random assignment to treatment and control is not feasible.



HIGHLIGHT

Project Spotlight: IRI Beacon
(<https://www.iribeaconproject.org/>)

The Beacon Project's interventions are informed through rigorous public opinion and media monitoring research, which is used to equip members of the Beacon Network with the tools and data to conduct in-depth analysis of malign narratives and disinformation campaigns. In 2015, the Beacon Project developed **>versus<**, a media monitoring tool used by in-house experts and media monitors across Europe to track malign narratives and disinformation campaigns in the online media space, analyze their dynamics, and how they are discussed online.

https://www.ndi.org/sites/default/files/NDI_Social%20Media%20Monitoring%20Guide%20ADJUSTED.pdf

To the extent analysts can collect network data over time, network analysis can also inform program monitoring and evaluation.

Data collection tools for network analysis depend on the nature of the network generally, and the network platform specifically. Network analysis can be conducted on offline networks where researchers have the capacity to collect data using standard face-to-face, telephone, computer-assisted, or SMS survey techniques. In these cases, researchers have mapped offline community networks using survey instruments that ask respondents to list individuals or organizations that are particularly influential, or whom they might approach for a particular task. Researchers can then map networks by aggregating and coding responses from all community respondents. In this way, researchers might determine which influential individuals in a community might be nodes for the dissemination of information, particularly in contexts where people rely largely on family and friends for news or information.

However, depending on APIs and terms of service, digital platforms such as social media can reduce the costs of network data collection. With dedicated tools, including social network analysis software, researchers can analyze and visualize relationships between users, including content engagement, following relationships, and liking or sharing

https://www.ndi.org/sites/default/files/NDI_Social%20Media%20Monitoring%20Guide%20ADJUSTED.pdf

These tools can provide practitioners with an understanding of the structure of online networks, and in conjunction with content analysis tools, how network structure interacts with particular kinds of content.



HIGHLIGHT

Tool Spotlight: [IFES/NDI VAWIE-Online Social Media Analysis Tool](https://www.ifes.org/publications/violence-against-women-elections-online-social-media-analysis-tool)

<https://www.ifes.org/publications/violence-against-women-elections-online-social-media-analysis-tool>

Information and Communications Technologies (ICTs) have created new vehicles for violence against women in elections (VAWIE), which are compounded by the anonymity and scale that online media platforms provide. A new tool from

the United States Agency for International Development (USAID), International Foundation for Electoral Systems (IFES), and National Democratic Institute (NDI) offers an adaptable method to measure the gendered aspects of online abuse and understand the drivers of this violence. The VAWIE-Online Social Media Analysis Tool can be used by actors from across a range of professions who are concerned by hateful and violent speech online and are motivated to end it.



HIGHLIGHT

Program/Tool Spotlight: NDI Data Analytics for Social Media Monitoring

NDI seeks to empower partners to leverage technology to strengthen democracy. This means harnessing technology's potential to promote information integrity and help build inclusive democracies; while also mitigating the harm posed by disinformation, online influence campaigns, hate speech, harassment and violence.

For that reason, NDI developed, "[Data Analytics for Social Media Monitoring](https://www.ndi.org/publications/data-analytics-social-media-monitoring) (<https://www.ndi.org/publications/data-analytics-social-media-monitoring>)," a guide for democracy activists and researchers.

This new guide is designed to help democracy practitioners better understand social media trends, content, data, and

networks. By sharing lessons learned and best practices from across our global network, we hope to empower our partners to make democracy work online by helping them:

- Collaborate with local, national, or international partners;
- Understand different methods of data collection;
- Make the best use of mapping and data visualization;
- Analyze the online ecosystem;
- Detect malicious or manipulated content and its source;
- Understand available tools for all aspects of social media monitoring; and
- Know how to respond with data, methods, research, and more through social media.



HIGHLIGHT

Program Spotlight: [Detecting Digital Fingerprints: Tracing Chinese Disinformation in Taiwan.](https://graphika.com/reports/detecting-digital-fingerprints-tracing-chinese-disinformation-in-taiwan/)

(<https://graphika.com/reports/detecting-digital-fingerprints-tracing-chinese-disinformation-in-taiwan/>)

In June 2019, with the 2018 local elections as a point of reference, Graphika, Institute for the Future's (ITF) Digital Intelligence

Lab, and the International Republican Institute (IRI) embarked on a research project to comprehensively study the online information environment in the lead up to, during, and in the aftermath of Taiwan's January 2020 elections, with an awareness of the 2018 precedents and an eye for potential similar incidents throughout this election cycle. Graphika and DigIntel monitored and collected data from Facebook and Twitter, and investigated leads on several other social media platforms, including Instagram, LINE, PTT, and YouTube. IRI supported several Taiwanese organizations who archived and analyzed data from content farms and the island's most popular social media platforms. The research team visited Taiwan regularly, including during the election, to speak with civil society leaders, academics, journalists, technology companies, government officials, legislators, the Central Election Commission, and political parties. The goal was to understand the online disinformation tactics, vectors, and narratives used during a political event of critical importance to Beijing's strategic interests. By investing in the organizations investigating and combating Chinese-language disinformation and CCP influence operations, they hoped to increase the capacity of the global disinformation research community to track and expose this emerging threat to information and democratic integrity.

RESEARCH TOOLS FOR UNDERSTANDING

DISINFORMATION

5. RECOMMENDATIONS (/TOPICS/SURVEYS/5- RECOMMENDATIONS)

- Develop research questions first, research designs second, and data collection methods and instruments third. To answer the questions that are most relevant for the context and program, research design and data collection methods should be selected to answer questions that are most important for the program measurement needs. Committing to a research method or data collection method before scoping your research question will limit what can be answered.
- In the implementation phase, consider a pilot-test-scale model for program activities. Using one or more of the outlined research approaches, workshop content on small groups of respondents, and use pilot data to refine more promising content before deploying activities to a larger set of beneficiaries.
- Protect personally identifiable information (PII). All of the data collection methods described in this section, from interviews, to surveys, to network data and social media analytics, can collect information on intimate and private personal characteristics, including demographic data, attitudes, beliefs, and willingness to engage in political action. Regardless of the selected methodology, researchers should make every attempt to secure informed consent to participate in research, and should take care to secure and de-identify personal data.
- Consider partnerships with research organizations, university labs, or individual academic researchers, who may have a comparative advantage in designing and implementing complex research designs, and who may have an interest in studying the effects of counter-disinformation programs.





Strengthening Democracy
Through Partnership

ABOUT (/ABOUT)

INTERVENTIONS (/INTERVENTIONS)

TOPICS (/TOPICS)

COPYRIGHT © 2023 COUNTERING DISINFORMATION